

The Journal of  
**APPLIED PSYCHOLOGY**

EDITED BY

**JAMES P. PORTER**  
Clark University

**WILLIAM F. BOOK**  
Indiana University

And a Board of Co-operating Editors

---

**VOLUME V**

1921

---

**WORCESTER, MASS.**

**FLORENCE CHANDLER, Publisher**

Reprinted with the permission of the original publishers.

**JOHNSON REPRINT CORPORATION**

**KRAUS REPRINT CORPORATION**

# The Journal of Applied Psychology

## CO-OPERATING EDITORS

W. V. BINGHAM, Carnegie Institute of Technology; JUNE E. DOWNEY, University of Wyoming; ROBERT H. GAULT, Northwestern University; WILLIAM HEALY, Baker Foundation, Juvenile Court, Boston; H. L. HOLLINGWORTH, Columbia University; JOSEPH JASTROW, University of Wisconsin; HERBERT S. LANGFELD, Harvard University; JAMES H. LEUBA, Bryn Mawr College; HENRY C. MCCOMAS, Princeton University; WALTER D. SCOTT, Northwestern University; C. E. SEASHORE, University of Iowa; DANIEL STARCH, University of Wisconsin; EDWARD K. STRONG, JR., George Peabody College for Teachers; L. M. TERMAN, Leland Stanford Jr. University; E. L. THORNDIKE, Columbia University; J. E. WALLACE WALLIN, Psycho-Educational Clinic, St. Louis; G. M. WHIPPLE, University of Michigan; HELEN THOMPSON WOOLLEY, Vocation Bureau, Cincinnati.

The scope of the JOURNAL OF APPLIED PSYCHOLOGY will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeal. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis. (c) The influence of general environmental conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or musical instruments, sports, etc.

---

## TABLE OF CONTENTS

---

L. DEWEY ANDERSON	
Estimating intelligence by means of printed photographs.	152-155
LEWIS O. ANDERSON	
A preliminary report of an experimental analysis of causes of stuttering . . . . .	340-349
ADA HART ARLITT	
On the need for caution in establishing race norms . . . . .	179-183
CHARLES E. BEDAUX	
The Bedaux unit principle of industrial measurement . . . . .	119-126
M. A. BILLS	
Methods for the selection of comptometer operators and stenographers . . . . .	275-283
A test for use in the selection of stenographers . . . . .	373-377
ELSIE OSCHKIN BREGMAN	
A study in industrial psychology—tests for special abilities	127-151
A. S. BRUDACHER	
The teaching personality quotient . . . . .	39-45
STEPHEN S. COLVIN	
Educational guidance and tests in college . . . . .	32-38
JESSIE C. FENTON and LEWIS M. TERMAN	
Preliminary report on a gifted juvenile author . . . . .	163-178
RAYMOND O. FILTER	
An experimental study of character traits . . . . .	297-317
MAX FREYD	
A test series for journalistic aptitude . . . . .	46-56
THOMAS R. GARTH	
White, Indian and Negro work curves . . . . .	14-25
The results of some tests on full and mixed blood Indians	359-372
ESTHER L. GATEWOOD	
An experiment in the use of music in an architectural drafting room . . . . .	350-358
A. R. GILLILAND and H. T. MOORE	
The measurement of aggressiveness . . . . .	97-118
C. F. HANSEN and M. J. REAM	
The predictive value of short intelligence tests . . . . .	184-186
WILLIAM HEALY	
Pictorial completion test II . . . . .	225-239
D. S. HERRICK	
A comparison of Brahman and Pancharna children in South India with each other and with American children by means of the Goddard form board . . . . .	253-260
HARRY D. KITSON	
Minor studies in the psychology of advertising from the Psychological Laboratory of Indiana University . . . . .	5-13
Industrial psychology in Europe . . . . .	287-290
F. KUHLMANN	
The results of repeated mental re-examinations of feeble-minded over a period of ten years . . . . .	195-224

## CONTENTS

DONALD A. LAIRD	
Paper and pencil research in the college classroom	378-381
HENRY C. LINK	
A new journal of practical psychology. A translation of its introductory article	85-88
GARRY C. MYERS	
Control of conduct by suggestion: An experiment in americanization	26-31
H. T. MOORE and A. R. GILLILAND	
The measurement of aggressiveness	97-118
L. W. PRESSEY and S. L. PRESSEY	
Further data with regard to sex differences	78-84
S. L. PRESSEY	
The problem of the unselected group in the standardization of tests	64-71
A preliminary investigation of general prognosis, i.e., "General Intelligence"	72-77
M. J. REAM	
A statistical method for incomplete order of merit ratings	261-266
M. J. REAM and C. F. HANSEN	
The predictive value of short intelligence tests	184-186
W. T. ROOR	
Two cases showing marked change in I. Q.	156-158
G. M. RUCH	
A preliminary study of the correlations between estimates of volitional traits and the results from the Downey "Will-Profile"	159-162
LEWIS M. Terman	
The status of applied psychology in the United States	1-4
LEWIS M. Terman and JESSIE C. FENTON	
Preliminary report on a gifted juvenile author	163-178
MORRIS S. VITELES	
Tests in industry	57-63
FRANK WATTS	
The construction of tests for the discovery of vocational fitness	240-252
ANNA KELLMAN WHITCHURCH	
Psychological norms among university freshmen	318-339
NOTES	95-96, 192-193, 285-291, 381
REVIEWS	89-94, 185-191, 292-295 382-386



# THE JOURNAL OF APPLIED PSYCHOLOGY

VOL. V

MARCH, 1921

No. 1

## THE STATUS OF APPLIED PSYCHOLOGY IN THE UNITED STATES

LEWIS M. TERMAN

In the August (1920) number of the *Psychological Bulletin* Dr. Boring, secretary of the American Psychological Association, summarizes statistics on the research fields in which 393 members of the association report themselves to be working. The statistics include all the members of the association except a few who failed to report on this item. A member who reported two or more fields was accredited fractionally to each of the fields, which accounts for the fractions in the figures given below. Of the 393 reporting, 53 are carrying on no research. The remaining 340 are classified as follows in Table VI of Dr. Boring's article:

FIELD OF RESEARCH	NUMBER ACCREDITED	
<i>I. Non-psychological Fields:</i>		
1. Philosophy.....	15.7	
2. Aesthetics.....	1.7	
3. Anthropology.....	2.6	
	<hr/>	
Total.....	20,	or 5.9%
<i>II. Chiefly Pure Psychology:</i>		
4. General psychology.....	8.7	
5. Experimental psychology.....	81.6	
6. Theoretical psychology.....	37.4	
7. Animal psychology.....	15.5	
8. Psychology and neurology.....	4.2	
	<hr/>	
Total.....	147.4,	or 43.4%

FIELD OF RESEARCH		NUMBER ACCREDITED— <i>Continued</i>	
<i>III. Social and Industrial Psychology:</i>			
9. Social psychology.....	6.1		
10. "Applied" psychology.....	10.7		
11. Industrial psychology.....	7.4		
	<hr/>		
Total.....	24.2,	or	7.1%
<i>IV. Chiefly Education:</i>			
12. Tests.....	51.2		
13. Educational psychology.....	43.9		
14. Education (including experimental pedagogy).....	13.2		
15. Clinical psychology.....	13.7		
16. Statistical method.....	1.2		
	<hr/>		
Total.....	126.2,	or	37.1%
<i>V. Psychopathology:</i>			
17. Abnormal psychology.....	7.2		
18. Psychiatry.....	2.5		
19. Psychopathology.....	7.0		
	<hr/>		
Total.....	16.7,	or	4.9%
<i>VI. Miscellaneous</i>			
	<hr/>		
	5.	or	1.5%
Grand Total.....	339.5	or	99.9%

Any classification like that attempted by Dr. Boring encounters certain difficulties which naturally prevent the results from being wholly unambiguous. I believe, however, that his data throw interesting light on the amount of research which is being carried on in "applied" as distinguished from "pure" psychology.

Omitting groups I and VI (non-psychological and miscellaneous fields of research) the above table shows 147.4 psychologists accredited to the field designated as "chiefly pure psychology," 126.2 to the field designated as "chiefly education," 24.2 to "social and industrial psychology," and 16.7 to "psychopathology." The last three groups may be thrown together as representing the total field of applied psychology. If these figures could be taken at their face value we would have the following division of American psychologists into "pure" and "applied."

Group II, pure.....147.4, or 47%  
 Groups III, IV and V, "applied".....167.1, or 53%

The matter, however, is not so simple. Lines of demarcation between research fields are hard to draw and my classification of the various fields listed by Dr. Boring into "pure" and "applied" is not wholly satisfactory. The following sources of error in particular should be mentioned:

(1) It is certain that not all of item 8 ("psychology and neurology") can rightly be accredited to "pure" psychology. Let us assume that half of the 4.2 individuals concerned should be accredited to "pure" psychology and the other half to neurology. This reduces "pure" psychology to 145.3. Moreover, since neurology may reasonably be classified as a non-psychological field, our 315 classified psychological workers are now reduced to 312.9.

(2) Doubtless some will object to classifying social psychology wholly in the "applied" group. Again we may compromise by placing half with "pure" and half with "applied" psychology. This raises "pure" psychology from 145.3 to 148.35, and reduced "applied" psychology from 167.1 to 164.05.

(3) Item 14 is classified by Dr. Boring as belonging entirely with the non-psychological subjects. However, since the item includes "experimental pedagogy" it is reasonable to suppose that a fairly large proportion of the individuals concerned could justifiably be grouped under "educational psychology" or "tests." Accrediting half to applied psychology and half to the non-psychological fields, the number in "applied" psychology is now reduced to 157.45 and the total number engaged in psychological research is reduced from 312.9 to 306.3.

On the basis of the above corrections the matter now stands as follows for the 306 members of the American Psychological Association here classified as engaged in research in the fields of "pure" and "applied" psychology:

"Pure" psychology .....	148.35 or 48.5%
"Applied" psychology .....	157.4 or 51.5%

From the above figures, it appears that considerably more than half of the psychological research which is being carried on by members of the American Psychological Association (which includes practically all the psychologists of the United States) falls in one or another of the fields of applied psychology. The true situation is probably even more favorable to applied psychology if, as the secretary of the association suggests, there is a greater tendency for a member to report a research field in which he formerly worked but which he is gradually leaving than a new field into which his interests

are carrying him. To the extent that such a tendency exists, the amount of research being done in the older and more established lines of psychology will of course be exaggerated by the figures. At any rate it is clear that considerably more than half the effort devoted to psychological research in the United States is in the applied fields.

This situation has developed almost wholly since 1900, and chiefly in the last decade. As the shift to applied psychology still continues, the figures will doubtless be even more striking a decade hence.

Is this trend unfavorable to psychology as a science? The more conservative will probably so interpret it, but I do not believe that the history of science warrants this view. One has only to recall how such sciences as physiology, bacteriology, chemistry, physics, and biology have been stimulated by human needs to realize that a science stands to gain rather than lose, even as pure science, when it enters the lists in the service of mankind. It is unlikely that psychology will prove an exception to the rule.

MINOR STUDIES IN THE PSYCHOLOGY OF  
ADVERTISING. FROM THE PSYCHOLOG-  
ICAL LABORATORY OF INDIANA  
UNIVERSITY

---

By HARRY DEXTER KITSON

---

*Introduction: The Use of the Historical Method in the Inves-  
tigation of Problems of Advertising*

Every time an advertiser composes an advertisement he faces questions which he would like to have answered. He longs for accurate measurements which would enable him to determine the probable value of various elements in the advertisement.

Obviously the time to make these measurements is after the advertisement is completed and sent on its mission. In seeking the answer to any question regarding advertising procedure, the most reliable way is to make actual trial of a method and keep track of returns; to try out another and keep track of returns. By successive trials of different methods and by comparison of returns from each, to arrive at a conclusion showing which method is most effective.

This scientific "investigation of returns," however desirable it may be, is many times not feasible. The returns from many forms of advertising, such as bill boards and car cards, do not come in measurable quantities. Moreover, the components of an advertisement are so numerous that it is difficult to isolate the several items and discover what part of the returns is due to each. Finally the expense of experimental campaigns is very great. Altogether, despite its desirability the method can hardly be employed in wholesale fashion.

A substitute has been proposed in the "laboratory method," according to which conditions are arranged as nearly as possible like the conditions of the market; subjects are chosen typical of the prospective readers of the advertisement; and measures are obtained of their reactions. In this manner various devices may be tried out before being used on a large scale. This method seems to promise good results; it may unquestionably throw light upon many questions, such as legibility of various type-faces, appropriateness of various colors

and shades. With many problems, however, it can hardly be used effectively owing perhaps to the artificial atmosphere of the laboratory.

But we have not exhausted the possibilities of scientific measurement in advertising. In case neither of the above methods is adaptable, we may secure light from still another direction. Our aim, be it remembered, is to discover with scientific accuracy the most effective way to do a thing before we proceed to do it. If we cannot accomplish this in the market or in the laboratory we may turn to the experiences of other advertisers who have faced our problem, and by observing their solutions we may govern our procedure.

How discover their experiences? To ask them would evoke contradictions and opinions of varying degrees of reliability. Furthermore, since we seek scientific formulations of our facts, we must have figures instead of opinions. How shall we reduce the experiences of advertisers to numerical terms?

The answer is, Go to actual advertisements. As the repositories of advertising experience they picture the trend of progress and provide guide posts showing the location of pitfalls and pointing out the directions in which success has been attained. Such an approach to the problems of advertising is called the "historical method." It is very simple to describe and operate, though it naturally requires the same care and caution that scientific method ever demands. In employing it we must first state our problem in the form of a question; then go to the files of newspapers and magazines such as exist in almost every library of parts. Among the advertisements contained therein we must search for the answer to our question; then we may tabulate our results and express them graphically.

In employing this method we make use of several assumptions:

First, that the practises which have persisted over a period of years have probably been of value. We judge that those firms which did not follow these practises did not make enough money to enable them to continue to advertise, or that those firms which did continue to advertise discovered the difference between valueless and valuable practises and adopted the latter.

We must not assume that advertisers made these changes as a result of conscious discoveries. They worked mainly by trial and error. As we shall see in graphs to be presented later, they discovered the usefulness of various practises only after long-extended periods of experimentation.

We may regard modern advertising practises, then, as products of unconscious evolution. In the struggle for existence in the world of publicity, certain features have survived because of certain psychological elements of strength and fitness. They exemplify a sort of economic "survival of the fittest."

A second assumption we make in utilizing the historical method is that if advertisers had studied the practises of their predecessors they might have avoided some errors and eliminated much wastage. And we assume that by surveying the practises of the past and by discovering which ones have survived and which have failed, we may modify our practises and find a shorter and more economical road to effectiveness in advertising.

We repeat that we are not proposing the historical method as a formula for the solution of all advertising problems. We advocate it merely as one form of scientific investigation which will enable us to supplement and corroborate information secured by other means.

We shall now offer the results of a series of investigations which will demonstrate the utility of the method, and will reveal facts about advertising practise of practical value to present-day advertisers.

### I. *Amount and Rate of Increase in the Use of Full-page Advertisements*

As one evidence of the growth of advertising, people often point to the increase in the use of full-page advertisements. They fail to give their contention its greatest possible weight, however, because they do not present supporting figures. The

TABLE I

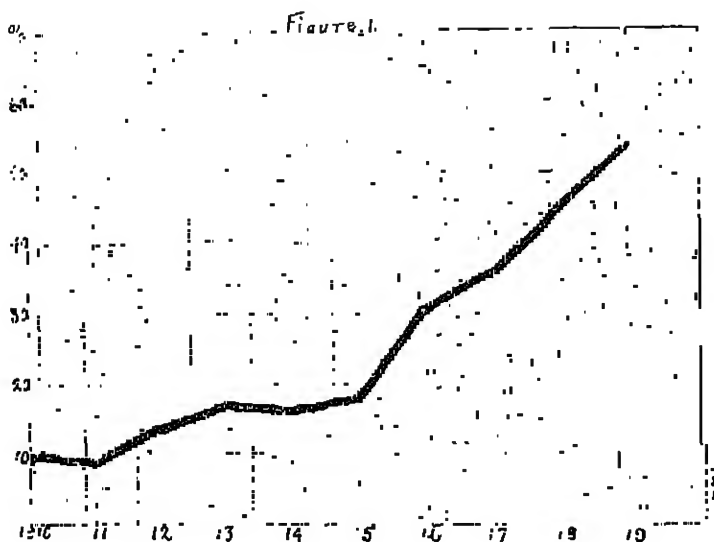
SHOWING THE NUMBER OF FULL-PAGE ADVERTISEMENTS (EXCLUSIVE OF COVER) IN THE LITERARY DIGEST FROM 1910 TO 1919

Year	Pages Devoted to Advertising	Full-page Advertisements	Per Cent
1910	1190	117	10
1911	1188	108	9
1912	1274	182	14
1913	1468	264	17
1914	1564	252	16
1915	1769	325	18
1916	2116	670	31
1917	2716	972	36
1918	2597	1162	45
*1919	2121	1155	54

\*6 months.

writer undertook to supply these by investigating advertising records for ten years back in one periodical—*The Literary Digest*. The ratio which the full-page advertisements bear to the total number of pages devoted to advertising was computed. (A double-page advertisement was counted as two singles.) Results are shown in Table I and Figure 1.

From these results we may state positively that the number of full-page advertisements has increased five-fold during the past ten years. Though these figures may not serve precisely



for all periodicals, still they are derived from a study of a representative general magazine, and they are surely indicative of the trend of development.

Our figures give further light upon the growth of advertising in general. As seen in the table, the total number of pages devoted to advertising in this periodical more than doubled during the past decade. The gross number decreased in 1918, due probably to exigencies of war economies. It should be noted, however, that the ratio of full-page advertisements to this total did not decrease; it even increased. Altogether, we have here numerical confirmation of the statement that full-page advertising has increased, and strong indication that advertisers consider it a paying investment.



## II. *Amount and Rate of Increase in the Height of Headlines in Full-page Advertisements*

Starch presents (page 40, Advertising) a table showing the increase in the height of headlines during the period 1870-1910, in which he shows that the height increased from 6.6 millimeters (approximately 24-point) to 12.4 millimeters (approximately 48-point). The writer undertook to ascertain if there had been any further increase since 1910. It was impossible to use the same magazine as those used by Starch (*Scribner's* and *Harper's*) as the advertising sections had been cut from the magazines in binding for the university library. *Collier's* was studied instead, with results as shown in Table II.

TABLE II

SHOWING AVERAGE HEIGHT OF HEADLINES IN FULL-PAGE ADVERTISEMENTS IN *COLLIER'S* FOR YEARS 1911-1920 INCLUSIVE

Year	MM.
1911	13.40
1912	12.77
1913	12.25
1914	12.15
1915	13.31
1916	12.76
1917	13.20
1918	12.52
1919	13.67
*1920	13.57
*First four months.	

If we accept the usage in this representative general magazine as indicative of the trend in advertising practise, we may conclude from these figures that there has been little change in the height of headlines in full-page advertisements during the past ten years—less than half a millimeter between the height in 1911 and that in 1920. Moreover the average height for the decade is 12.9—but one-half millimeter greater than that reported by Starch for the year 1910. From this we conclude that Starch's statement, "the maximum size of display type for a magazine page has now been reached," is substantially correct; and that on the average, 48-point type has proved of greatest effectiveness.

## III. *The Use of Borders with Full-page Advertisements*

In an investigation of this problem Starch found (p. 83-4, Advertising) that the use of borders with full-page advertisements in "standard magazines" increased markedly from 1880

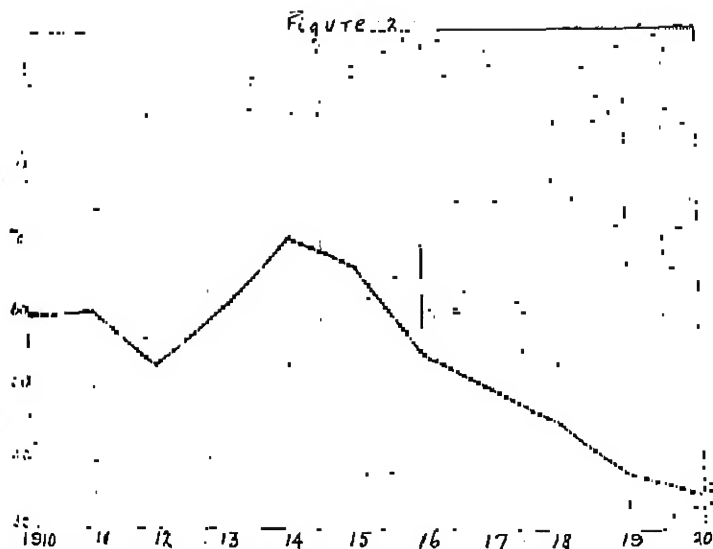
to 1910. The percentages of full-page advertisements having borders during the years 1880, 1890, 1900, 1910 were 6, 38, 82 and 77 respectively. Starch concludes, "there has been a remarkable increase from the time when large advertisements rarely had a border until to-day (1914) nearly every one has a border." He opines that "the slight fluctuation for 1910 is probably an accidental fluctuation."

TABLE III

SHOWING THE NUMBER OF FULL-PAGE ADVERTISEMENTS WITH BORDERS IN COLLIER'S WEEKLY FROM 1910 TO 1920

Year	Full-Page Advs.	Borders	Per Cent
1910	284	171	60
1911	310	188	61
1912	285	151	53
1913	355	199	56
1914	336	239	71
1915	302	203	67
1916	465	255	55
1917	467	232	50
1918	413	189	45
1919	574	217	38
*1920	478	168	35

\*20 issues.



In order to secure exact information covering the practise during the decade 1910-20, the writer conducted an investigation of the advertisements appearing in *Collier's Weekly*, with results as shown in Table III and Figure 2.

Figure 3.



These figures show that instead of a tendency toward increased use of borders with full-page advertisements, there is a tendency toward decreased use of same. Our results indicate that the drop in Starch's figures of 1910 was really significant; that the percentage is now back where it was in 1890. The tendency is shown graphically in Figure 3 based upon the data gathered by Starch, and brought down to date by the writer's investigation.

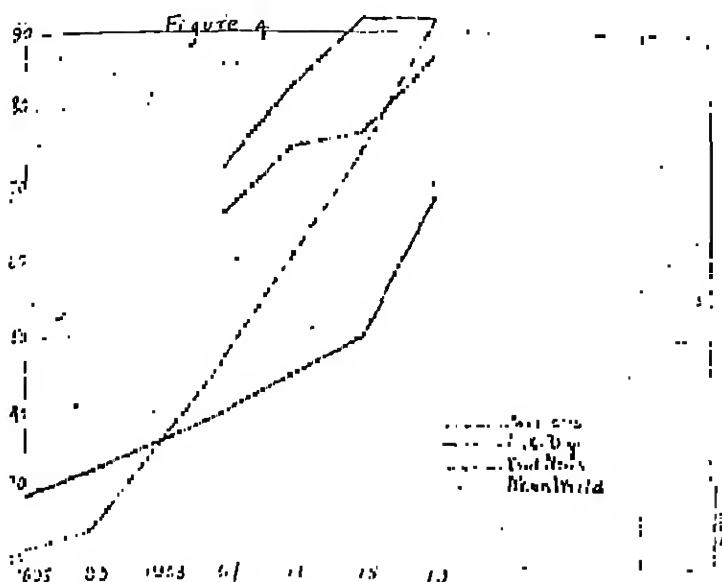
Whether the history of borders represents an evolution directed by the varying effectiveness of different methods under the conditions of the various eras, or whether it represents merely changes in fashion is difficult to determine. Undoubtedly there are fashions in advertising practise; particularly in matters that touch art-forms as closely as do borders. Fashions, we know, move in cycles. Hence, it may be that this curve represents merely the rounding out of a cycle which will later revolve in favor of borders again. Whatever is the true explanation, we may confidently conclude that among progressive advertisers of to-day, the full-page advertisement without border seems to be most in favor.

#### IV. *Amount and Rate of Increase in the Use of Illustrations*

In the investigation of this question four mediums were examined—two weekly magazines: *Collier's* and *The Literary Digest*; and two daily newspapers: *The Indianapolis News*, a typical metropolitan paper, and the *Bloomington (Ind.) World*, a typical small-town sheet. Measurements were made at four-year intervals from the issues 1895-1919. (The years 1903 and 1911 are missing in the case of the newspapers.)

The results, as shown in Figure 4 enable us to state that the relative number of illustrations has tripled during the past twenty-five years. Whereas in the earlier period scarcely thirty advertisements in a hundred were illustrated, to-day ninety in a hundred are so embellished. Though the small-size newspaper has not yet reached this percentage, both the magazines studied have done so, and the larger newspaper almost reaches it.

Using our conception of the "survival of the fittest" in advertising, we may well conclude that the use of illustrations in advertising has paid. Our figures present a striking argument to the advocates of purely "argumentative" copy and suggest that the advertiser who wishes to profit by the experience of the past will do well to use illustrations in abundance.



### *Summary*

In this series of investigations we have presented figures bearing upon some of the important problems which every copy-writer faces: utility of full-page advertisements; height of type for headlines; borders for full-page advertisements; value of illustrations. For light we have gone to the records of advertising usage, past and present. We have reasoned that the practises which have persistently recurred have probably been valuable, and that we may save ourselves some errors by adopting them. We do not champion the "historical method" as the best method of scientific investigation in advertising but we hope that we have demonstrated its adaptability to certain types of problem, and its serviceableness to advertising practise.

NOTE:—Assistance in gathering the data for these investigations was rendered by Miss Grace Morris, and Messrs. Robert Wiles, Robert Reed, Grady Stubbs, Carlyle Purden, Robert Norman, Brantley Burcham, S. L. Crawley, students in the Psychology of Advertising at Indiana University.

## WHITE, INDIAN AND NEGRO WORK CURVES

By THOMAS R. GARTH, University of Texas

The investigator working in the field of racial psychology is looking either for similarities or differences, or perhaps for both. On the one hand he notes the statement of Lester F. Ward, who stood actively for the equalities, or similarities:

"It is not therefore proved that intellectual equality, which can be safely predicated of all classes in the white race, in the yellow race, or in the black race, each taken by itself, cannot also be predicated of all races taken together, and it is still more clear that there is no race and no class of human beings who are incapable of assimilating the social achievement of mankind and of profitably employing the social heritage." *Applied Sociology*, Lester F. Ward, 1906, p. 110. And on the other hand he reads from Professor Giddings, who emphasizes the differences:

"It is sometimes said that we ought not to assert that the lower races have not the capacity for social evolution, because we do not know what they could do if they had opportunity. They have been in existence, however, much longer than the European races, and have accomplished immeasurably less. We are, therefore, warranted in saying that they have not the same inherent abilities." *Principles of Sociology*, F. H. Giddings, 1896, p. 328.

Woodworth's epoch making work in 1904 at the Louisiana Purchase Exposition (*Science*, 1910, No. 31, N. S.) tends to level the differences between races to almost a vanishing point in so far as more elementary processes are concerned. However, he said it was possible that the form board performances indicated differences of a racial character, but called this a mere "crumb" for the seeker of racial differences. Bruner made experiments on the hearing of primitive peoples at the same time as Woodworth and in his conclusion to the monograph by the title "The Hearing of Primitive Peoples" (*Archives of Psychology*, No. 11, July, 1908), says: "The one fact standing out most prominently as a result of these measurements is the clearly evident superiority of whites over all other races, both in the keenness and in the range of hearing sense. The evidence is so clear and

striking as to silence effectually the contention that the hearing function, inasmuch as it is of relatively less utility in the pursuits attending modern social conditions than those surrounding the life of the savage, has deteriorated and is degenerating. . . . Other things being equal those individuals or races possessing the greatest complexity and variety of reactions to elements in their respective environments likewise will be gifted with keener and more acute sensory mechanisms," pp. 111-112. On the other hand he says, on page 101 *ibid.*, "Furthermore, if we compare this group of native Indians with those who have been in attendance at the U. S. Government Schools,"—both of these groups were included in the races compared with the whites—"it will be noted as striking that the more intelligent Indians—those having been subjected to the influences of civilization—have a better auditory acuity than do those who have been closest to nature and a natural life. So far as our American Indians of the plains are concerned, therefore, it cannot be averred that their senses deteriorate with increased contact with civilization." We are inclined to conclude from this that the differences here mentioned among the Indians are in training and not so much in innate differences.

Likewise some of the more recent investigators in race psychology place the emphasis on the differences. Mayo and Ferguson working on the Negro are inclined to believe the differences between Whites and Negroes are rather large as based on their investigations (*Archives of Psy.*, No. 28, 1913, Mayo; *Archives of Psych.*, No. 36, 1916, Ferguson), while Pyle claims that his results show specific differences between the same races (*School and Society*, Vol. 1, 1915, p. 360). In 1917 Miss Sunne tested Negro and white children in New Orleans with the Binet-Simon Scale and Yerkes' Point Scale. The conclusion reached by this investigator is that variations in individual capacities is probably as great an influence in causing differences as either race or sex. (*Jour. of App. Psy.*, No. 1, 1917, p. 83.) Rowe tested White and Indian children and found rather decided differences as indicated by the Binet-Simon Scale. (*Ped. Sem.*, No. 21, 1914, p. 456.)

The present writer in an article "Racial Differences in Mental Fatigue" (*Jour. of Applied Psy.*, June, 1920), calls attention to certain apparent differences in work curves, i.e., the tendency for Indians to fall away, or "fatigue," less than the white or negro groups who worked at the same task continuously, i.e., work in simple addition. This was true for

the groups of younger and older children of the races compared in the two categories of *attempts* and *accuracies*.

Popular opinion holds to the side of great racial differences. Sometimes this persuasion is so extreme that it is called, in those cases, race prejudice. In some circles race prejudice is looked upon as innate, but it is more likely that it is a tradition with a rather long line of descent having once had a great eugenic value for keeping pure a racial stock. However, it would be well for the popular mind to take caution for we do not know but that experimental psychology may eventually render the term "inferior races" as innocuous as it seems it has the long honored term "the weaker vessel" as used in connection with sex differences.

The elements in a study of racial mental similarities or differences must be these: (1) Two so-called races  $R_1$  and  $R_2$ , (2) an equal amount of educational opportunity,  $E$ , which should include social pressure and racial patterns of thought, and (3) psychological tests,  $D$ , within the grasp of both racial groups. We should have as a result of our experiment  $R_1$ ,  $E$   $D$  equal to, greater than, or less than  $R_2$ ,  $E$   $D$ . In this experiment the only unknown elements should be  $R_1$  and  $R_2$ . If  $E$  could be made equal the experiment could be worked.

This element of educational opportunity-nurture, is the one causing most of the trouble in racial psychology as an uncontrollable element. It does not offer quite so much difficulty in the study of sex differences, yet it is there only in smaller degree than in racial differences, and as it is controlled the "sex differences" tend to disappear. Since this element of education, or nurture, cannot be eliminated it would be safer to take for comparison such racial groups as have had as nearly the same educational opportunity as is possible having any disparity of this sort well in mind when we interpret the results of the experiment. Having done this, we first take the complete distributions on the scale of measurement for the groups as statements of the true facts of the case, race for race. We then combine these distributions into a total distribution of accomplishment of all the races taken together to see if we have multimodal effects. Should we find these effects we may conclude that we have evidence of types, or racial types, and there should in this case be one mode for each racial group. But should the combined distribution for the several racial groups reveal only one mode we may conclude that the test reveals no types—no real racial differences but rather similarities. (See Thorndike, *Educational Psy.*, Vol. III, p. 322, p. 347.



We may then further examine the distributed data for overlapping in the several groups and measure that, using one group distribution as a basis for the comparison.

This is what we have endeavored to do with the work curves of Whites, Indians, and Negroes. The work curves are not given here, but will be found in the article by the writer referred to above.

The question is this: Are there separate and typical work curves for Whites, for Indians, and for Negroes?

Under the title "Work Curves," *Journal of Educational Psychology*, May-June, 1919, the writer showed that for White children, one type of work curves exists for a specified sort of work (single column addition) for certain age groups, under certain conditions. The same method of handling the data is employed here.

Third and Fourth Grade Subjects---Attempted Performance.

TABLE I  
PER CENT OF WORK DONE IN FIRST SIX MINUTES

Per Cent	W.	Ind.	Neg.	T.
1-5	1	..	..	1
6-10	3	1	0	4
11-15	18	6	5	29
16-20	119	35	17	171
21-25	181	43	58	282
26-30	37	7	17	61
31-35	7	0	1	8
36-40	0	0	1	1
41-45	0	..	..	..
46-50	0	..	..	..
51-55	1	..	..	..
56-60	0	..	..	..
61-65	0	..	..	..
66-70	0	..	..	..
71-75	0	..	..	..
76-80	1	..	..	1
Total	368	92	99	559

TABLE II  
PER CENT DONE IN FIRST 14 MINUTES

Per. Cent	W.	Ind.	Neg.	T.
16-20	1	..	..	1
21-25	0	..	..	..
26-30	1	..	..	1
31-35	3	0	1	4
36-40	9	3	2	14
41-45	37	20	7	64
46-50	55	39	36	230
51-55	21	19	35	176

TABLE II—*Continued*

Per. Cent	W.	Ind.	Neg.	T.
56-60	29	7	13	49
61-65	8	1	2	11
66-70	1	1	1	3
71-75	0	2	0	2
76-80	1	..	1	2
81-85	1	..	..	1
86-90	0	..	..	...
91-95	0	..	..	...
96-100	1	..	..	1
Total	368	92	99	559

TABLE III

PER CENT DONE IN LAST SIX MINUTES				
Per Cent	W.	Ind.	Neg.	T.
0	3	0	0	3
1-5	..	0	1	1
6-10	1	1	2	4
11-15	24	4	6	34
16-20	114	26	44	184
21-25	196	48	43	287
26-30	22	10	3	35
31-35	6	2	0	8
36-40	2	1	0	3
Total	368	92	99	559

TABLE IV

3rd and 4th Grade—Accurate Performance.

PER CENT DONE IN FIRST SIX MINUTES

Per Cent	W.	Ind.	Neg.	T.
0	27	5	0	32
1-5	1	0	1	2
6-10	14	1	2	17
11-15	29	11	10	50
16-20	86	30	18	134
21-25	95	27	24	146
26-30	59	7	27	93
31-35	23	5	6	34
36-40	10	3	4	17
41-45	4	1	2	7
46-50	10	1	2	13
51-55	2	0	0	2
56-60	1	0	2	3
61-65	0	0	..	...
66-70	4	0	1	5
71-75	0	..	..	...
76-80	0	..	..	...
81-85	0	..	..	...
86-90	0	..	..	...
91-95	0	..	..	...
96-100	3	1	..	4
Total	368	92	99	559

TABLE V  
PER CENT DONE IN FIRST FOURTEEN MINUTES

Per Cent.	W.	Ind.	Neg.	T.
0	7	2	0	9
1-5	0	0	0	...
6-10	0	0	0	...
11-15	0	0	0	...
16-20	4	0	0	4
21-25	3	0	1	4
26-30	5	1	0	6
31-35	10	1	2	13
36-40	19	7	4	30
41-45	44	15	11	70
46-50	81	26	27	134
51-55	72	20	17	109
56-60	46	11	20	77
61-65	24	3	7	34
66-70	24	1	5	30
71-75	7	3	2	12
76-80	6	0	0	6
81-85	2	0	1	3
86-90	1	0	0	1
91-95	1	0	1	2
96-100	12	2	1	15
Total	368	92	99	559

TABLE VI  
PER CENT DONE IN LAST SIX MINUTES

Per Cent	W.	Ind.	Neg.	T.
0	37	8	4	49
1-5	3	6	2	5
6-10	29	5	5	39
11-15	52	12	19	83
16-20	89	17	29	135
21-25	80	30	23	133
26-30	47	11	13	71
31-35	18	4	3	25
36-40	4	2	1	7
41-45	2	2	..	4
46-50	1	1	..	2
51-55	2	..	..	2
56-60	3	..	..	3
61-65	0	..	..	0
66-70	1	..	..	1
Total	368	92	99	559

## 7th and 8th Grade Subjects—Attempted Performance.

TABLE VII

## PER CENT DONE IN FIRST SIX MINUTES

Per Cent	W	Ind.	Neg.	T.
1-5	..	..	..	1
6-10	16	2	2	20
11-15	228	75	16	319
16-20	97	19	14	130
21-25	2	1	2	5
Total	343	98	34	475

TABLE VIII

## PER CENT DONE IN FIRST TWENTY-TWO MINUTE!

Per Cent	W.	Ind.	Neg.	T.
32-36	4	1	..	5
37-41	3	..	..	3
42-46	18	6	3	27
47-51	99	38	9	146
52-56	171	43	19	233
57-61	39	10	2	51
62-66	8	..	1	9
67-71	0	..	..	..
72-76	1	..	..	1
Total	343	98	34	475

## 7th and 8th Grade—Attempted.

TABLE IX

## PER CENT DONE IN LAST SIX MINUTES

Per Cent	W.	Ind.	Neg.	T.
0	1	..	..	1
1-5	1	..	..	1
6-10	17	1	3	21
11-15	232	73	21	326
16-20	87	22	10	119
21-25	2	1	..	3
26-30	1	1	..	2
31-35	2	..	..	2
Total	343	98	34	475

7th and 8th Grade—Accurate Performance.

TABLE X

PER CENT DONE IN FIRST SIX MINUTES				
Per Cent	W.	Ind.	Neg.	T.
0	2	1	..	3
1-5	3	..	..	3
6-10	37	14	1	52
11-15	170	47	10	227
16-20	105	25	17	147
21-25	21	9	4	34
26-30	1	2	1	4
31-35	2	..	..	2
36-40	1	..	..	1
41-45	0	..	1	1
46-50	1	..	..	1
Total	343	98	34	475

7th and 8th Accurate Performance.

TABLE XI

PER CENT DONE IN FIRST TWENTY-TWO MINUTES

Per Cent	W.	Ind.	Neg.	T.
31-35	..	1	..	1
36-40	3	..	..	3
41-45	11	3	1	15
46-50	75	31	10	116
51-55	110	49	10	169
56-60	100	13	10	123
61-66	26	1	1	28
66-70	11	..	1	12
71-75	1	..	1	2
76-80	2	..	..	2
81-85	1	..	..	1
86-90	3	..	..	3
Total	343	98	34	475

TABLE XII

PER CENT DONE IN LAST SIX MINUTES

Per Cent	W.	Ind.	Neg.	T.
0	4	..	2	6
1-5	6	..	..	6
6-10	49	20	7	76
11-15	205	44	14	263
16-20	73	25	11	109
21-25	6	5	..	11
26-30	..	4	..	4
Total	343	98	34	475

TABLE XIII

## THIRD AND FOURTH GRADE ATTEMPTS

<i>(a) First Six Minutes</i>			
	Whites	Indians	Negroes
Median.....	21.81%	22.08%	23.46%
% reaching median of Whites		54%	78%
<i>(b) First Half of Total Time</i>			
	Whites	Indians	Negroes
Median.....	50.28%	49.88%	51.89%
% reaching median of Whites		43%	65%
<i>(c) Last Six Minutes</i>			
	Whites	Indians	Negroes
Median.....	21.59%	22.69%	20.13%
% reaching median of Whites		68%	37%

TABLE XIV

## THIRD AND FOURTH GRADE ACCURATES

<i>(a) First Six Minutes</i>			
	Whites	Indians	Negroes
Median.....	22.41%	20.95%	24.00%
% reaching median of Whites		42%	68%
<i>(b) First Half of Total Time</i>			
	Whites	Indians	Negroes
Median.....	51.52%	50.25%	52.75%
% reaching median of Whites		44%	55%
<i>(c) Last Six Minutes</i>			
	Whites	Indians	Negroes
Median.....	19.82%	21.64%	19.28%
% reaching median of Whites		69%	48%

TABLE XV

## SEVENTH AND EIGHTH GRADES—ATTEMPTS

<i>(a) First Six Minutes</i>			
	Whites	Indians	Negroes
Median.....	14.91%	14.66%	15.95%
% reaching median of Whites		46%	64%
<i>(b) First Twenty-two Minutes</i>			
	Whites	Indians	Negroes
Median.....	53.02%	52.33%	53.56%
% reaching median of Whites		43%	53%
<i>(c) Last Six Minutes</i>			
	Whites	Indians	Negroes
Median.....	14.99%	14.61%	14.50%
% reaching median of Whites		46%	44%

TABLE XVI

## SEVENTH AND EIGHTH GRADES—ACCURATES

<i>(a) First Six Minutes</i>			
	Whites	Indians	Negroes
Median.....	13.10%	15.07%	16.67%
% reaching median of Whites		49%	73%
<i>(b) First Twenty-two Minutes</i>			
	Whites	Indians	Negroes
Median.....	54.67%	52.36%	51.50%
% reaching median of Whites		26%	49%
<i>(c) Last Six Minutes</i>			
	Whites	Indians	Negroes
Median.....	14.10%	14.67%	13.62%
% reaching median of Whites		57%	44%

The younger groups worked twenty-eight minutes, the older groups worked forty-two minutes and all racial groups worked under the same conditions. There were two categories—attempts and accurates.

Since distributions for all points in the curve are out of the question the curves were tapped in three places and three distributions were made of these tapings: (1) A distribution of what the individuals of a group did in the first six minutes, (2) a distribution of what had been done by the time about one half of the total working period had elapsed, and (3) a distribution of what the individuals did in the last six minutes. The tables presented herewith give the racial group distributions—Whites, Indians, and Negroes separately, and the three racial groups combined into a total in each instance.

It will be seen that wherever we tap the total curve (made of the three racial groups of performers) the distribution for same has a single mode, whether it be the case of distribution at the beginning of the curve, the middle of the curve or the end of it. Though each racial curve has its idiosyncrasies as found in that curve itself this idiosyncrasy is lost in the total distribution. And these idiosyncrasies by the present test of the data show no great departure from the common curve. This single modality holds for all the curves—younger and older groups in both attempts and accuracy, and although the curves have been tapped in only three places, the beginning, the middle, and the end, these are very vital points and it is likely that tapping at other points would reveal the same situation—a uni-modal distribution curve. In consequence this handling of the data is but a support of the claim that

none of these curves are racial in a typical sense but each contributes toward a common *human* work curve, and to that extent stands for similarities rather than differences.

We now in the second place take the expanded data and examine it for overlapping using as a basis of comparison the medians of the White groups.

Granting that these three groups really belong in one large group, the question still remains: How does one sub-group compare with another? See Tables XIII, XIV, XV, and XVI. It will be seen that there are only three instances in which the overlapping of the White and Indian distributions is as much as 60% and above, whereas there are six cases of overlapping above that amount when the Whites and Negroes are compared. The most significant instance is where at the beginning 78% of the younger Negroes attained the median of the younger Whites in attempts. Only 37% of them attained the median of the Whites for attempts at the end, whereas 68% of the Indians of the younger group attained this same White median at this point. In fact, as we have shown elsewhere, the Indians make a good showing when compared with the Whites and Negroes with respect to the last six minutes of work in all instances excepting where only 46% of the older group attain the median of the Whites of the older group in the attempts category, for this part of the common curve, and the Negroes consistently fall below the median performance of the Whites and Indians during the last six minutes of work.

The figures representing what was done by the time about half of the period of the total working time had elapsed are interesting. If one works absolutely at a regular rate he should, of course, have done 50% of his work by the time half of his total working period was up. By running over the above mentioned tables (Tables XIII, XIV, XV, and XVI), one may observe that consistently more Indians were inclined to do less of their total work than the median White did, and that more Negroes were inclined to do more of their work than the median White by the fourteenth minute, for the younger workers, and the twenty-second minute for the older workers. This means then that slightly more than half of the Indians were even as deliberate, or more so, in their work as the median White, and that half of the Negroes were inclined to be less deliberate than the median White—with one exception only: as in the accurates category of the older Negroes where they did just the same part of their work as did the Whites.



These measures of overlapping are significant as far as they go and are to be taken into very careful consideration before one takes a stand for a belief in racial work curves of Whites, Indians, and Negroes or in just a common *human* work curve. It is a fact that they incline one to see differences and we may add that mixture of blood in the Indians and in the Negroes may have prevented these differences from being more apparent.

## CONTROL OF CONDUCT BY SUGGESTION: AN EXPERIMENT IN AMERICANIZATION

By GARRY C. MYERS, Cleveland School of Education, Cleveland, Ohio

When the course of study and the texts\* which are now used in the Americanization Schools of the Army were developed at the First Recruit Educational Center at Camp Upton there was a conscious attempt to make as the warp and woof of the lessons a propaganda for that school, for more education by the soldier and his home folks, for the Army, for the United States government, and for America and American ideals. It was assumed that all suggestions of proper habits and virtues are most effective when camouflaged and especially when the learner unwittingly becomes a teacher of those duties and ideals which it is desired he shall get. The learner is not told that he should sleep with his windows open, that he should brush his teeth and clothes, stand erect, sit erect, take care of his appearance and respect the great men and flag of our country. Instead the learner becomes the reader or the writer of a letter to a friend in which he tells that he has developed these desirable traits, attitudes and virtues and suggests indirectly or directly to this friend that the latter could do the same and attain the same virtues and attributes. Throughout, the learner suggests the very virtues to himself which he seems to be suggesting to his friends. A few sample lessons follow:

Co. D., Recruit Educational Center,  
Camp Upton, N. Y.  
December 7, 1920.

Dear Mother:

It is pretty cold some morning when we must jump out at Reveille but we must move so fast then that we do not mind it. A soldier does not feel the cold much because he sleeps with his windows open. When I came to the army I was sure I would take cold if I slept where the windows were open, but I soon learned that fresh air is the best cure for a cold. I have never felt so well in all my life as I feel

---

\* Army Lessons in English by Capt. Garry C. Myers, War Department, 1920.

now. Almost every soldier says the same. The army is the place to be. I wish Zebediah would join the army. Then he would learn to clean his teeth, to stand erect, sit erect and walk erect, and to keep his elbows out of his pockets.

If you see Nellie tell her I still like cake.

Jeremiah Butler.

From the next lesson is the following:

"You know that Theodore Roosevelt had four boys, and they were all soldiers. One boy, whose name was Quentin, gave up his life for his country in the Great War. We honor his memory and we honor the memory of his father. Every good American is ready to do the same as Quentin Roosevelt did. A long time ago, when our country was fighting for freedom a great American soldier by the name of Nathan Hale said as his last words, 'I only regret that I have but one life to give for my country.' Don't you think that was wonderful?"

Moreover, those first lessons of the course leading up to letter-writing aim to "sell" the course and the school to the learner; to suggest to him that he wants to learn and why he wants to learn. The first sentence he reads is "I want to learn English" and the fifth is "I want to learn to write a letter home." In his eighth lesson he reads as a sample a first letter to his sweetheart:

Co. B, Recruit Educational Center,  
Camp Upton, N. Y.

January 8, 1920.

Dear Nellie:

I got your letter and the sergeant read it for me. The next letter I am going to read myself for I am learning to read and write English. I am writing this letter myself. It is my first letter. I hope you like it. I like the army very well. I want you to see me drill. I can drill very well. We drill every day.

Jim Jones.

In his next lesson, which is a letter to his mother, he reads:

Co. C, Recruit Educational Center,  
Camp Upton, N. Y.

February 1, 1920.

My dear Mother:

This is my first letter home. I am going to write you a letter every week. I got a long letter from Nellie. She said she saw you. The sergeant read it for me but her next

letter I am going to read myself. I want to read a good book. It is a hard job to learn to read and write but a good soldier likes a hard job. I have a good captain. He is a friend to every soldier.

Robert Brown.

In these letters he suggests to himself, that he is going to read the next letter from his sweetheart, that he is going to read a good book, that he is doing well, a big job and that his captain is his friend. In succeeding lessons he reads and writes from sample letters the following suggestions:

"Every good soldier knows how to handle a gun."

"I am sorry that I did not go to school at home, I did not think then that I could learn, but I am sure now that I can learn. I can write letters to Nellie and I can read her letters to me."

"I have read a few letters by Theodore Roosevelt to his boys and I want to learn to read some speeches by Woodrow Wilson, President of the United States."

In the third grade, his second lesson is a letter to Nellie from which he reads:

"I have learned to brush my teeth, to keep my head and shoulders up, and to keep my uniform neat and clean. I wish I could get my teeth to look as clean and as pretty as your teeth."

It should be remembered that these are not mere statements but personal expressions to one, Nellie. The dramatization carries over very well, so well that the average soldier enthusiastically plays the part in "writing" these letters.

The non-English soldier who reads the following from a letter to "Dear Bozo" speaks with pride:

"When we go to New York we can see men and women coming into a new land just as we came five years ago, and we know how they feel. Of course, we want them to like America as we like America. We want them to know that the cleanest, best and most useful American is the soldier. Therefore when I go to New York I want to look my best."

Careful records have been gathered which indicate that the men take over these letters by phrases, sentences and often in their entirety, when they write home. Most of these men have limited mental content; so they are glad to grasp at almost any word, phrase or sentence suggested. Undoubtedly they write home much which they at first did not believe and over which they were not always enthusiastic. Furthermore, those at home, because they look upon these soldiers, their near friends, with esteem, catch up the suggestions.

As good evidence that the lessons read by the soldier tended to express themselves in what that soldier actually did write home there are given below copies of several letters or excerpts from them. It should be noted that practically every sentiment expressed in these examples had appeared once or several times in the lesson-letters of the reading texts.

*Some Actual Letters-Home*

"Dear Mother:

Yesterday I received the socks which you sent me. I can write letters now as you see. I am sure you will find me looking better then when I left. I am a better man I look much better and feel much better then I did before I enlisted in the army and I know much more. I am for the army."

"I always keep my clothes clean and have to keep my teeth clean. I have to brush my teeth. I received the socks that you sent me."

Dear Sister Ethel

I have ben in the army now about three monts an when I first came in the army I Dident rely know what our flag ment. But it means much more than red whit an blue cloth Every stich an every thread means many American soldiers who suffered died an bled. For thear country an thear country is our country an I am proud to be a soldier They did not suffer an die for themselves but for thear country for their farther mother an brothers sisters an sweethart. An I have sure learned to honor our flag it means more than you could think Tell brother I wish he would join the Army."\*

Dear Mother.

I have not heard from you for long time. I hope you are not ill. I have been in the American army three months and I like it very well. When I came in the army I could not read or write English, but now I can read and write a little it has been three years that time when I cane in this country and am in school now to want learn good English "

Dear Mother

I am in school today I like to go to school We go to school every morning and every evening. I would like to see you. I may be home before long on furlough. I feel much better since I joined the army. I like the army fine. We have

---

\*FOOT-NOTE.—This is a faithful reproduction in paraphrase of a lesson-letter on the American flag.

plenty of good clothes to wear. we go to the moving picture show every night. We have plenty of good books to read and we have a good time."

"dear mother yesterday from a walk i learned to write letters I am go to write every week I am a soldier I am go to read a book from abraham lincoln as I will close"

"I am. this army. 6 months and this camp. Upton i learn. som. end. go to scoold. every. day. learned read. and. write. now. he can write letter myself home. before and. go to my sergent read my letter. now i can. read. myself. I write letter and read were well."

"My dear Mother.

This is my frist letter to you. I am surprise you that I am writing. to let you now that I am Well. I hope your our Sampn. Mother I have A good Captain he is a friend to every Soldier. I go to School every day and I now how to read. and writer. our. Captain learn. os how to handle A gun. and we like to drill. every Soldier likes to handle a gun. good luck."

"I am for the Army because oncle Sam dew his best for me I like you see me now I wight 150 lbs. my Capt said I contined aet so much I wight 200 lbs the Co. B, his the best Co in the Camp Upton I got a very good Capt."

"My dear Mother.

this is my frost letter I write you, and I hope to see you soon. dear Mother I am filled much better now and much strong I am sure you will find me looking better then when I left home.


Dear Mother I like the army very will we have plenty eat and good clothes"

"before I Came to the army I could not write my name every menth I get a good money I have good clothes plenty to eat I do not pay for them"

Another avenue of control of conduct of the illiterate soldier at Camp Upton was in the public speaking course. The soldier in this class presumably was getting training to help him qualify for recruiting duty. This very purpose suggested to him that he must make speeches magnifying the virtues of the school and of the army. He proceeded therefore in impromptu speeches to tell of what the school had done for him. Consequently there developed a strong rivalry among the members of this class (public speaking being voluntary) to make the most embellished speech about the very things

which were suggested in the lesson letters. Some who had scarcely begun to read and write extended their imagination so as to tell, as if it were true, of how well they had learned to read and write English since coming to that school. By suggesting to a number of the old time "chronic-kickers" of the Camp that they would make good speakers, many of them became strong competitors for first place among the boosters of the school, their company, the camp and the army. They kept at this boosting for several class periods until they really began to believe what they were saying and to act accordingly. Thus by their own suggestions they became effectual missionaries for morale.

Little doubt is there that in education one of the most effectual means for controlling conduct is by suggestion.



## EDUCATIONAL GUIDANCE AND TESTS IN COLLEGE\*

By STEPHEN S. COLVIN, Director of the School of Education, Brown University, Providence, Rhode Island

At the beginning of the academic year 1919-20 Brown University inaugurated a system of educational advice and direction for its students, the details of which are still in the making. The most essential features so far put into operation are the following:

(1) On the first day of the college year the first-year men are given the Thorndike Mental Test for College Freshmen.

(2) A few weeks later these men are given the so-called Brown Psychological Test. The Brown University Test is composed of two parts; the first requiring twenty-five minutes for administration and the second, thirty minutes. Separate elements of the test are: completion, vocabulary, opposites, mixed relations, facts-and-conclusions, and arithmetic. This test requires about a third of the time of the Thorndike Test, but shows results almost identical, correlating with the Thorndike Test by a Pearson coefficient around .90 and frequently showing a closer agreement with the Thorndike Test than do separate parts of the Thorndike Test itself. The reason for giving both tests is to check up the occasional contradictory results that are found and to discover their causes.

(3) Early in the year each new man is required to fill out a personnel card which through a series of questions seeks to discover:

- (a) The purpose the student has in coming to college.
- (b) His academic interests while in high school.
- (c) His activities in various sports and out-door life.
- (d) His reading interests.
- (e) His vocational aptitudes and interests.

In addition to answering questions concerning these matters the student is required to give an estimate of himself in regard to twenty-five qualities relating to intelligence, habits of study, character, temperament, social interests and health.

---

\* This paper was read at Atlantic City, Feb. 25, 1921, before the National Vocational Guidance Association.



(4) Later in the year each student is interviewed by a member of the Committee on Educational Advice and Direction, who after consultation with the student fills out a second personnel card, which is filed with the card previously made out by the student. On this second card are recorded the scores made in the psychological tests and all academic grades that the student may have received, in addition to other items of interest such as whether the student entered on condition, whether he has received any academic honors, whether he has been warned for poor work or is on probation. Also there is added any information the student may be able to give concerning his secondary school career; the academic and vocational interests and plans that he may have; his habits of study, with particular reference to the place, time, technique and amount of study; his outside interests and distractions, including work for self-support, extra-curricula college activities, and amusements and recreations.

In addition, from the interview with the student, his advisor discovers and records the following facts, as far as possible: (a) the birth place of the student and of his parents; (b) the number of children in the family and the order of birth of the student; (c) the occupation of the father and his approximate income; (d) the occupation and schooling of brothers and sisters with particular reference to the grade that they have completed in school and their age at the time of finishing this grade; (e) the age at which the student graduated from high school together with any other information concerning his school career.

(5) Early in the year a blank is sent to the principal of the school in which the student prepared for college, asking information on the following points: (a) the rank of the student in his class, with the number of members in the class; (b) the courses in which the student did his best and his poorest work; (c) the extra-curricula activities of the student; (d) his interests; (e) his qualities of initiative and leadership. In addition, the principal is asked to give an estimate in regard to qualities of intelligence, character and social adaptabilities possessed by the student.

(6) During the first half of the freshman year the students are required to attend a course of orientation lectures given by various administrative officers and instructors. These lectures are designed to inform the student concerning important details of college life, ideals and methods of study, hygiene and personal habits, and the aims and content of general fields of knowledge represented in the curriculum.

(7) The students are again interviewed by their advisors at the end of their sophomore and junior years. At the end of the sophomore year each student is required, under direction, to select certain courses and groups of studies in terms of a definite scholarly interest or a life career. These studies are to satisfy the requirement for concentration, not, however, in terms of departments of knowledge (as was previously the case) but in terms of interests and aptitudes.

(8) In addition to these regular interviews, students are required to consult with the chairman of the Committee whenever particular reasons for such consultation arise, such as poor or unusually excellent college work, on disparity between psychological scores and college records. One particular function of the Committee is to stimulate students of high ability and attainment to do work worthy of their capacities. For this reason, the direction of reading for final honors, open only to students of unusual intelligence and academic achievements, is in the hands of the Committee.

(9) Finally all students in the University are urged to consult with the Chairman of the Committee during his general office hours in regard to any problems connected with their college work or their life careers which may be important for them.

It will be seen from this account that psychological tests play by no means the entire part in the work of guiding the student. They, however, are extremely useful in various ways and are frequently of great service in clearing up doubtful points in regard to the students' abilities, aptitudes and achievements. They have been found particularly valuable in the following ways:

1. They throw some light on the question as to whether a student is better suited for a professional career or for business pursuits. They cannot do this, however, in any but a most general manner since the range of intelligence among college men is narrow and since practically all who remain for any length of time in college receive scores that place them somewhere in the professional occupational intelligence group as determined by the army intelligence ratings. Brown men obtain a median intelligence score that is somewhat higher than that of civil engineers and medical officers and somewhat lower than that of army chaplains and engineering officers. Occasionally a student receives a score that would place him on the occupational level of clerical and semi-professional workers, but never so low as that of the skilled laborer. At

present the general practice of the advisors is to urge men of high intelligence ratings to enter some one of the professions, provided that in addition their character, interests and special capacities seem to qualify them for such a life career. This practice is based not only on the consideration of the abilities of the students but also on the needs of society. It is found that very few young men, even those of high ability and character, are planning on careers in medicine, law, teaching and the ministry. On the other hand, the large majority of students, as far as they have formed any definite purposes, are thinking of business, though in a rather hazy fashion.

2. The tests indicate to an extent the type of mind that a student possesses. For example,—a student who does much better in his college work than on his mental tests may be a slow but rather accurate thinker who learns less readily than some of his fellow students but who retains and assimilates what he once masters. There is some evidence to show that psychological examinations of the type of the Thorndike and Brown tests are unfair to men who are plodders but careful and accurate thinkers. Here is a specific case in point.

A student of the present junior class who appears to be a man of at least average ability and of more than average earnestness of purpose received a score in his psychological tests that indicated him to be of unsuitable intelligence to succeed in college work. His college grades, however, have been above the average. When questioned about his psychological tests he said that they took him by surprise and that the time limits bothered him. He finds that he can respond rapidly to something that he has memorized, but cannot think under pressure. The student was later retested with an alternative mental examination in which no time limit was fixed for the various sub-tests. The result was that he raised his score by thirteen points, confirming the diagnosis that he was of about median college intelligence and that he owed his success largely to careful and regular habits of study rather than to brilliancy or rapidity of thinking.

3. Intelligence tests make possible a distinction between character qualities and mental alertness. It is evident from the examination of Brown students that intelligence considered by itself is by no means a perfect measure of success or failure in college. Of course, intelligence is undoubtedly positively correlated with character, and an intelligence rating is never a rating for pure mental ability alone, but is also to a degree an index of volitional and other character qualities. But further than this, numerous cases have been discovered in which

the disagreement between intelligence scores and college grades is clearly due to qualities of character and temperament. Generally the disagreement is evidenced by students who receive high intelligence ratings and low college grades; sometimes the reverse is true. Examples follow:

A member of the present sophomore class scored high in all his mental examinations. He graduated from high school at the age of seventeen, standing in the upper tenth of his class. His college record, however, has been disappointing. He confesses that he is lazy in his mental habits and lacks determination and ambition to study.

In contrast to this student is another member of the same class, who is clearly somewhat below the average in intelligence but who is determined to succeed in his work and who is making good grades because of his will-to-do.

Some students succeed or fail in college because of emotional qualities. As an illustration of this fact the following case is in point: A member of the present sophomore class obtained a high rating in the Thorndike Test and an average rating in the Brown Test, the lower rating in the latter test being due probably to physical disability while taking this test—the student having fainted a short time before, something that he says occurs quite frequently. He graduated from high school at nineteen, having been out of school two years because of ill health. He has two sisters and one brother, all of whom are at least one year in advance of their normal school grade. There seems to be little doubt that the student's intelligence is above the average; his work in college, however, has so far been almost a complete failure. He is now on final probation and has a record of six E's (failures) out of a total of fifteen courses taken. He says that he is extremely nervous and that he cannot endure the strain of the ordinary college examinations. He got on well in high school because of the fact that his grades were so high that he was not obliged to take formal tests to pass his courses. In the formal college tests he "gets chills and is all-gone."

4. At times tests throw light on the home environment and educational equipment of the student. Sometimes students who score relatively low in their psychological tests do so because of the fact that they hear a foreign language spoken at home and for the most part think in a foreign language. This is particularly true of a group of Italian students now at Brown. Their college work is markedly better than what seems probable from the prognosis made on the basis of their mental scores.

In this connection it may be interesting to note that in a recent comparison of the results of mental tests given pupils in one grade in a Providence school the results of the Terman revision of the Binet Tests were found in general to agree rather closely with the results of the National Intelligence Tests as far as American children were concerned. In the case of children of Italian parentage, however, the mental ages were from one to three years lower in the National Tests than in the Terman Tests. The tentative conclusion is that the National Tests, largely linguistic in their nature, failed to reveal the real intelligence of the Italian children because of their unfavorable home environment as far as the speaking and reading of English is concerned.

5. Tests sometimes reveal the presence or lack of scholarly ambitions and proper education ideals. Students often take the attitude of treating their college work as a task to be performed rather than an opportunity for improvement. This is not true merely of those who are content to work only just enough to "get by," but this point of view is all too frequently found among men of high ability who can get really good grades on an amount of study often averaging not more than six hours a week. It is very clear that our higher institutions of learning, in common with our lower schools, are not challenging the abilities of their best students.

6. Tests often show whether it is desirable for a student to continue in college or to withdraw. Intelligence tests have not as yet been used at Brown to determine college entrance, except in a few exceptional cases. They frequently indicate, however, whether it is worth while to keep a man in college or not. If a student has an extremely low intelligence score, and if his grades are markedly below average, there is little use in keeping him on semester after semester. It is equally true that a student of ability who constantly fails is undesirable college material because of his lack of character and ideals.

7. The results of the tests frequently serve as incentive to students to do work up to the level of their mental abilities. It is the writer's practice to show to a student who is doing college work distinctly below his real ability his case in a graphic form. For this purpose a wall chart is made which indicates each student's position in the psychological ratings and also in his academic achievement. It evidently produces some impression on a man to say to him,—“Here you stand in the first third of your class in your psychological scores,

and yet you are here (twenty points lower) in your college rating. You are not living up to your possibilities. You know the facts; now it is up to you." I am confident that such an objective presentation carries vastly more weight than any good advice could possibly bring to bear. Facts are more impressive than general statements and sometimes produce results which no amount of exhortation could accomplish.

## THE TEACHING PERSONALITY QUOTIENT

By A. R. BRUDACHIER

It is merely stating the obvious to say that profound scholarship is not even presumptive evidence of teaching power. The converse is equally true, that teaching power does not necessarily betoken great scholarship. Our colleges have great teachers whose scholarship is mediocre and both colleges and schools have suffered from men of deep, broad and sound scholarship whose teaching was and is a travesty on their high profession.

In training teachers, persons are discovered who develop an intellectual acumen out of all proportion to their social and volitional power. These persons come to the end of their course with high ratings for scholarship, with a thorough and broad knowledge of professional technique, but when they attempt the teaching process they fail utterly to project either themselves or their subject into the class before them. We say of them that they are immature, that they lack social experience, that they lack personality. This judgment is tempered by our knowledge of numerous cases of a like sort, where greater maturity, wider contact with life, have developed a dynamic personality. And for this reason we are properly cautious in our judgment of all minus or merely colorless personalities among teacher recruits. We hesitate to declare that any specified person will never develop this teaching power; but in this hesitation lies the source of the continual stream of poor teachers, teacher failures, the derelicts of the profession.

This personal power is the unknown quantity in higher education but particularly in teacher training. The school of education addresses itself to the task of developing scholarship and to the inculcation of special methods in a *restricted* field. The college entrance requirements are defined wholly in terms of scholarship, the test being a test of the intellect, and that largely a test of retentive memory. That memory is a very small element in teaching power has not disturbed our equanimity; that intellect is only a fraction of the teacher's dynamics, has not affected our plans. We have made the humble confession that personality is a vague term, diffi-

cult of definition, a complex resultant of heredity, of environment, experience and training, and this humility covers a host of professional sins.

We have made no consistent effort to analyze the teaching personality, either for the purpose of eliminating from the ranks of teachers in training, those who have it not; nor yet for the purpose of developing and increasing and improving those phases of personality which are present at the outset in a small degree.

The intelligence test, so called, is the first hopeful sign that there is a way out. It gives promise that there is a means of testing personal power, of discovering personal qualities which rest in the emotions and the will either in whole or in part. And tests of the motives which lie behind human behavior, if they can be made, will shed bright light on personality and will bring into high relief the elements which are the bases of the teaching personality in particular.

The "intelligence quotient" is, true to its name, primarily a test of the intellect. Its superiority to the conventional educational tests lies in the fact that it tests not only *memory*, but the *power of perception*, the *association of ideas*, *reason*, the *power of judgment*, *comprehension*, *ingenuity*, etc. That is, while the conventional examination may call for any or all of the intellectual qualities enumerated, it is difficult and even impossible to say how much of the student's reaction is a pure memory product; but in the intelligence tests, the other qualities are momentarily in focus and memory is not. This is in itself a great gain. But for the analysis and the discovery of the teaching personality still other qualities must be brought into focus, qualities that are wholly absent from or are present only by implication in the intelligence test.

For the sake of conformity to familiar terminology, I propose to derive a "Teaching Personality Quotient." Our study at the New York State College for Teachers divides the matter into four phases, comparable roughly with, and including the Intelligence Quotient. These phases are as follows: A Personal History Quotient; an Intelligence Quotient; a Behavior Quotient; and a Special Information Quotient.

(1) The Personal History Quotient. The personal history of the prospective teacher is the first phase in point of time. It has many elements. Environment shapes personal characteristics. Nationality and the occupational history of the family are to be recorded. The cultural background of the family states the educational opportunities of parents and grandparents and indicates the character of the student's home



life. The high school record continues to record the student's scholastic achievements as in the past. A physical examination throws light on the student's physical and mental condition. There will be a careful record regarding the vocational, avocational and recreational interests so far as such interests have become definite. And the student's achievements will be tabulated—achievements in leadership, in organization, in carrying responsibility, in exhibiting courage, persistence, industry, etc. The *personal history quotient* will serve as the basis for professional guidance throughout the educating period. Merely to state the case for this quotient is to show its immense superiority over the present inadequate matriculation process, consisting as it does, almost exclusively of the scholastic record.

(2) The Intelligence Quotient. The intelligence test will be first of all confirmatory of the personal history. It will illuminate the scholastic record as contained in the high school certificate or in the entrance examination. It will be invaluable in checking our conventional judgments on scholastic attainment. But the intelligence quotient will do much more. It will furnish an index of intellectual power, which, taken in conjunction with the student's record of youthful achievement, will enable the college to understand shortcomings and to provide intelligent corrective measures for failures in specific things. For example, if the test discloses a relatively weak power of association, reason and judgment, the correct diagnosis may be followed by appropriate prescriptions. This fact gives greater importance to these elements in the intelligence quotient than to such elements as mental alertness and retentive memory which will probably not yield greatly to training and education. In this respect our use of the intelligence quotient differs radically from the use made by the army.

(3) The Behavior Quotient. It is not unusual for students to make violent changes in their social habits when they pass from school to college. Leadership demonstrated in the school, may not be in evidence in the larger social group. Avocational interests may be changed. Even vocational interests may be affected by the new environment. The persistence exhibited in the school, the purposefulness of the earlier time may not be apparent. These changes in behavior are insignificant in our present scheme because we make no effort to build the college life in consistent fashion on the life habits developed in the pre-college period. Viewed merely as an intellectual process, the college career has only incidental relationship to this body of habit which accompanies the entering Freshman.

But in a larger scheme of personality cultivation it is extremely important to insure continuity of development for such fundamental characteristics as were noted above, persistence, purposefulness, leadership, vocational and avocational interests. Unfortunately the college has not yet sensed its responsibility here, and has no machinery for the discovery of personal characteristics, for the systematic and continued development of well defined qualities, for the strengthening of qualities that appear weak or only nascent as yet. I believe this is an unscientific procedure in all higher education. Teacher training surely cannot attack its problem intelligently without this systematic study and continuous development of these simpler phases of personality.

The Behavior Quotient will set up this desirable basis of personality cultivation. It will rate the Freshman on those points which determine character, those qualities which give color to his behavior in the class room, among his associates, in all social relations, everywhere. The rating for behavior must of course be restricted to those qualities which can either be tested under controlled conditions, or which can be observed and recorded with accuracy. The Freshman will readily suspect the purpose of the test and yield the expected result because he detects the desire of the tester. This will vitiate the results of the test. Consequently, it will be highly important to select those qualities of conduct, emotional and volitional and intellectual, which lend themselves best to the purposes of the test. In fact, it will require long and patient endeavor, with a study of hundreds of cases, to establish reliable results. But of this I feel certain at the outset. A small list of qualities may be tested under strict laboratory conditions, and other qualities may be observed and tabulated during the college course.

(a) Behavior by Test. I propose the following list of qualities for the determination of the behavior quotient by test. Some of these may later have to be eliminated, others may be brought in as more desirable. Each constitutes an important part of the teaching personality.

(1) *Sympathy with and understanding of children.* The instinct of play is absent in some persons. The playful child annoys them. Naiveté is incomprehensible to these people. Fairy tales, childish imaginings are foolishness to them. The absence of this quality will be a veritable blindspot in the teaching personality.

(2) *Resourcefulness*. The ability to use ways and means that are not superficial or readily apparent, constitutes an indispensable element of teaching personality.

(3) *Directive ability*. The ability to determine the direction which a group of people or a class shall take in thought or action.

(4) *Self Control*. The ability to reserve judgment, to resist seductive suggestion, to control the emotions, to maintain equanimity in spite of provocation.

(5) *Tact*. This is a composite of many qualities but for purposes of testing is conceived to be the power of quick analysis of a social situation, a correct understanding of the requirements, and a rapid marshalling of forces to meet the situation. Tact brings the best solution with the least irritation to the situation. A rare, but an indispensable quality in the teaching personality.

(6) *Purpose*. A clear view of things as they ought to be and a definite desire to make them so.

(7) *Organizing ability*. The ability to adopt means to ends, whether the situation be mental or physical.

(8) *Speed of Reaction*. The ability to respond rapidly to social situations.

(9) *Aesthetic Sense*. A sense of beauty; good taste; pride in personal appearance, including neatness and cleanliness.

(10) *Humor*. A cheerful view of life; a right sense of proportion; the habit of correct self appraisal, understanding the relationship between dignity and self-conceit.

(11) *Pervasive attention*. The ability to give primary attention to one thing and at the same time attend in secondary degree on all objects within the sensory field. This is a prerequisite to class control.

There will probably be immediate agreement that the qualities named are fundamental if not indispensable in the teaching personality. But can they be tested? Our presumption is that social situations can be constructed and questions asked regarding the conduct recommended, which will clearly define the character of the student under test. For example:

"Of a number of recorded actions, which do *you* approve?"

"Given these or those conditions, what will *you* do?"

"In a specified and complicated situation, what solution do you offer?"

A faculty committee in the New York State College for Teachers is at work on the construction of tests. The plan

is to extend the method of the intelligence test so as to cover cases where intelligence, together with feeling and will, affect social situations in the form of concrete behavior. The list of qualities was selected from a large total; Dr. Fairchild's "Perfect Human Being Chart" for instance, was rich in suggestiveness; but the question, Can it be tested? reduced the list to the qualities named.

(b) Behavior-by Observation. There are qualities that can be observed but cannot be disclosed and appraised by test under controlled conditions. They are fully as important as any that have been named and will shed much light on the teaching personality. The behavior quotient cannot be determined with finality without them and so other means must be devised for their study and appraisal. This means lies in the systematic observations by those members of a faculty who have frequent opportunity for such study. This is the major officer in many institutions, he who gives the fundamental courses which form the major interest of the student; he who advises the student on all matters pertaining to electives and the satisfaction of the requirements leading to graduation and to a degree. This same officer, may also, in the school of education, give instruction in the special method offered in the major subject. His judgment is based on many meetings, after observing the student in many situations, and should command correspondingly great respect. Other instructors may have opportunity for such systematic observation.

The qualities to be submitted for this type of personality study are the following:

Enthusiasm	Adaptability
Industry	Loyalty
Responsibleness	Respectfulness
Cooperation	Poise
Aesthetic Sense	Independence

A special rating card will be provided and a definite set of values assigned.

(4) The Special Information Quotient. This consists of the student's record in his major and minor courses, his scholarship record generally, and his work in special method. This record can be compiled finally, only at the end of the college course and will be valuable in guiding the prospective teacher into a position but not in guiding him during his period of preparation. This phase of the study offers nothing different from the usual college scholarship record. The variation lies rather in the use that is to be made of this record. Hereto-

fore scholarship rating was the only definite judgment passed on students. By the proposed plan it will take its place in a larger scheme.

It is not time to propose final relative values for each of the quotients here proposed. Our investigation has not progressed far enough for that. But it is already clear that the plan will be helpful in passing judgment on the graduate who seeks recommendation to a teaching position. That the employing agency—school superintendent or trustee or college president, will find a judgment based on such a personality frame trustworthy, is also quite likely. That such employing agency would scan the personal history quotient and the behavior quotient with as much care as the special information quotient is at least probable.

The personality study here proposed seeks to do three specific things. (1) It endeavors to bring together under the four heads, all the evidence regarding the student's personal characteristics and to provide the information needed for professional guidance. (2) It seeks in extreme cases to enable the college or the school of education to advise individual students against preparing to become teachers, or against continuing such preparation already begun. (3) It seeks to provide an intelligent basis for genuine personality cultivation. By shifting the center of emphasis, even slightly from scholarship alone, by bringing in elements of character and character building, it will make of teacher training what it should be, a process of discovering and developing the dynamic personality. The chief value will lie, not in the elimination of the minus personality, but in training and developing the plus personality. By discovering in each individual student the best combination of personal qualities which he possesses, and by the nurture of this innate strength, the educating process may compensate his defects. For we all have personality negatives. It is the projective power of our plus characteristics, that wins success, that determines the size of the teaching personality quotient.

## A TEST SERIES FOR JOURNALISTIC ATTITUDE

By MAX FREYD, Research Fellow, Carnegie Institute of Technology<sup>1</sup>

The present study was begun as an attempt to devise and evaluate a series of psychological tests for use in vocational selection and guidance in the field of journalism; more specifically, to provide for college instructors in journalism a means of foretelling the probable success of their students as reporters. The advantages of such tests to both student and instructor need no great elaboration: they provide for the elimination of unfit students from classes in journalism, enabling the instructor to give more time and attention to those who have the requisite ability, and sparing the unfit student the years of effort to acquire information which will bring him no adequate return.

The criterion by which the tests will eventually be evaluated is success in newspaper reporting after graduation from college. They will be given to a sufficient number of undergraduates to insure a thorough trial. In technique this study resembles closely that of Dr. L. L. Thurstone's study of engineering aptitude: the tests will be given to journalism students, preferably in their first year at college, and the test scores will be compared with criteria of progress in college, and eventually with success in reporting after graduation.

The difficulties usually met with in such analyses were encountered in determining the traits of the successful reporter. Text-books of journalism are full of psychographic pictures of the profession. Münsterberg, in his "Vocation and Learning," listed what he thought were the qualifications for reporting. These analyses consist mainly of moral qualities whose presence in a remarkable amount might well be conceived as constituting a drawback to success. A study of reporters on the job, discussions with newspaper men, and a careful sifting of the literature on the subject, led me to the

---

<sup>1</sup> The material in this report was gathered at the University of Washington during the academic year 1919-1920. The writer is indebted to Dr. Stevenson Smith for advice throughout the study.

conclusion that the following traits, above others, were responsible for good reporting:

High degree of intelligence	Social ability
Broad range of information, especially on current events	"Nose for news"
Good memory	"Nerve"
Language ability	Keen interest in reporting

Tests were devised for all except social ability, "nerve," and interest in reporting. Recent refinements in tests for volitional and social traits and in the use of interest questionnaires make it possible to provide objective measures of all the items in the list in any future study along this line.

### 1. *The Tests*

The series originally used consisted of nine tests printed in booklet form for group testing and so arranged that only one test was visible at a time. Each test was timed, and the whole series could be given during the regular class period. A revised series is being used, at present, but all the results given herein were obtained with the original series.

The following description of the tests is accompanied by a statement of the purpose for which each test was devised. The tests, of course, measure a multitude of other characteristics besides the ones mentioned, but our assumption is that they measure the traits mentioned more efficiently than they do other traits.

Test 1 is intended to measure the complex ability known as "a nose for news," which means the ability to see an event immediately in terms of its news value and to recognize the items that would be of the greatest interest to the greatest number of readers of a newspaper. The term may also be taken to include the ability to know where to look for news. It is probably acquired largely through training, but the extent to which the student possesses the ability before his training may be indicative of his future performance in the vocation.

This test consists of ten pairs of statements, all of which refer to a tenement house fire. The subject is required to indicate by a check which of the items in each pair would be more important in a news story of the fire, assuming that he is a reporter covering the fire for a newspaper in the city in which it occurred. In constructing the test an effort was made to pair an item of major importance with one of minor importance. The following are samples:

- ( ) While responding to the alarm, a hose truck skidded into a telephone post, but was not damaged and proceeded to the fire.
- ( ) The tenement house was a three-story brick building, erected in 1880.
- ( ) At the time of its construction the building was one of the show places of the city.
- ( ) A crowd of several hundred people watched the fire.
- ( ) The firemen had difficulty in preventing the flames from spreading to neighboring buildings.
- ( ) The fire started at 6:30 a. m. and was not put out until 9:30 a. m.

In order to determine the "correct" answers, this test was sent to the editors of forty leading newspapers of the country and to a number of instructors in journalism. The items in the test were evaluated on the basis of the nineteen answers which were received. In most cases there was a large majority for one of the items in each pair.<sup>2</sup>

Test 2 has a purpose similar to that of Test 1, namely, to find out how well the student can solve problems in reporting without previous practice. The test consists of fourteen situations which might confront a reporter, to each of which three solutions are offered. The subject checks with a number

---

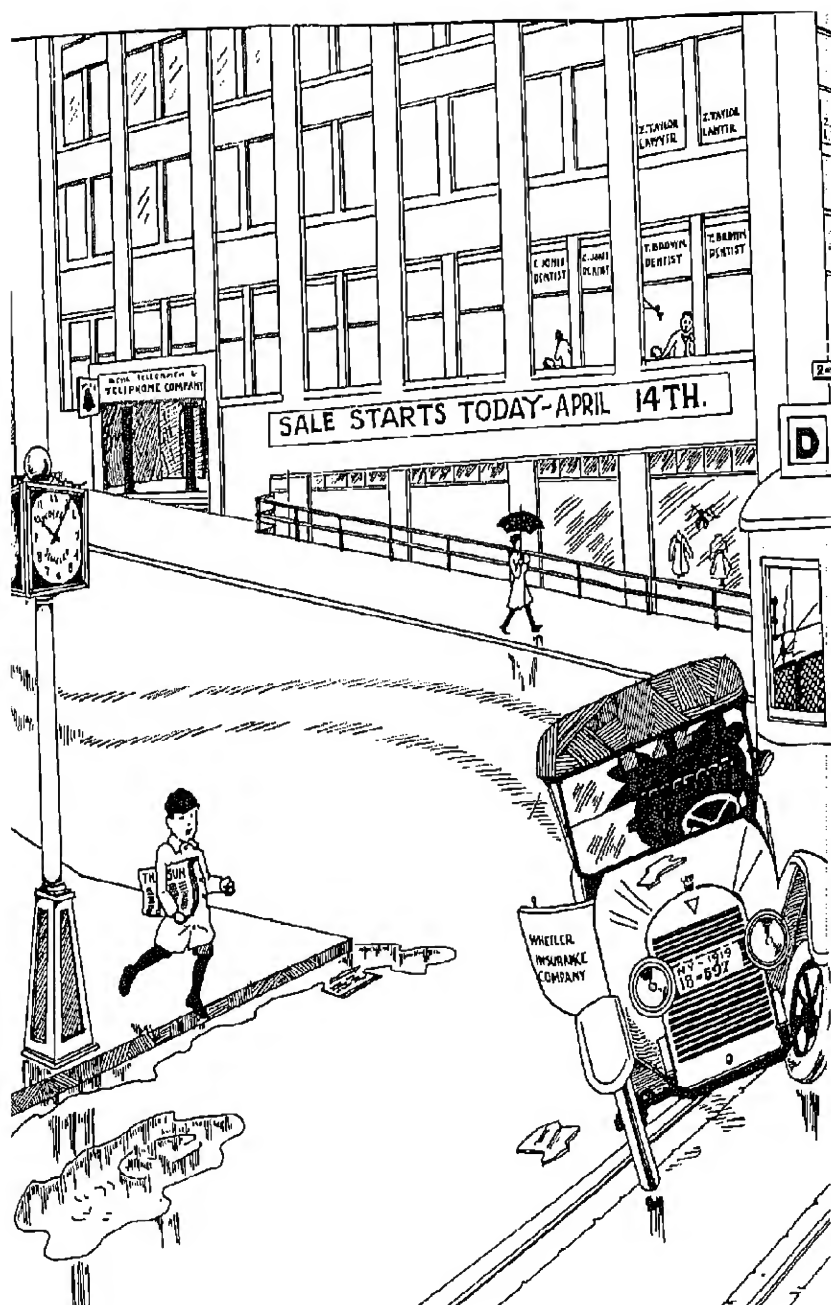
<sup>2</sup> In scoring this test, it is desirable that three factors be taken into consideration: (1) The value assigned to the right answer must bear some relation to the percentage of the total votes cast that it received; (2) Inaccuracy must be penalized, since there is a choice between only two answers; and (3) The difference in score between checking the right answer and checking the wrong answer in a pair must be proportional to the value assigned to the right answer.

Let  $x$  represent the total number of votes cast for both statements in a pair, and let  $y$  represent the number of votes cast for that statement which received the greater number of votes. By assigning values to the right answers in proportion to the ratio  $y/x$  the first factor is accounted for. Values were assigned to the right statements according to the following table:

$y/x$	Value assigned
.50 to .55	0
.56 to .65	1
.66 to .75	2
.76 to .85	3
.86 to .95	4
.96 to 1.00	5

In order to account for the second and third factors the same amount is subtracted from the individual's total score for checking a wrong answer as would have been added to it if he had checked the correct answer. Thus a difference in score between checking the right answer and checking the wrong one is always double the credit assigned to the right answer.







1 the solution which he considers the best of the three, and with a 2 the second best solution. Sample questions are:

If you were assigned to get the facts about a bank failure, what would you do first?

- ( ) Go to the bank and knock at the doors.
- ( ) Call the bank on the phone.
- ( ) Call up the office of the bank examiner.

What would you do if a great fire broke out a few hours after you had quit work for the day?

- ( ) Hunt up the owner of the property and interview him.
- ( ) Call up your office.
- ( ) Get all the data you could about the fire from direct observation.

What would you do if a prominent man gave you a copy of a speech he was about to deliver before a large audience, and while delivering it he made some extreme statements that did not appear in the written copy?

- ( ) Look him up and ask him if he meant those statements for publication.
- ( ) Leave those remarks out of your account, as probably being due to the excitement of the occasion, and not expressing his real point of view.
- ( ) Include those statements in your account, along with the written speech.

In order to evaluate the answers, this test was sent to the same group as the first test. In all except three of the problems there was a very large majority for one of the three answers. So many of the men overlooked registering their second choice that it was decided to score the test merely on the first choice.

One criticism of this test which is not without point is that the subject if actually confronted with the situation would not do what he indicated. It is probable that the same may be said of the editors who filled out the test. The test merely shows how well the student agrees with editors of long experience in solving theoretical problems.

Test 3 measures auditory memory for meaningful material. A short selection saturated with facts is read to the students, after which they are shown a printed copy of the selection in which blank lines have been substituted for many of the important facts. They are asked to fill in the gaps from memory. The test is scored by means of a stencil which is placed over the printed copy.

Tests 4, 5, 6, and 9 were intended to be measures of various phases of language ability. The following are taken from Test 4:

- |                  |   |
|------------------|---|
| ( ) learned      | Possessing information                            |
| ( ) educated     |   |
| ( ) informed     |   |
| ( ) narcotic     | Having the quality of causing insensibility or    |
| ( ) astringent   | stupor.   |
| ( ) sedative     |   |
| ( ) autonomy     | The power, right or condition of self-government. |
| ( ) freedom      |   |
| ( ) emancipation |   |

The subjects place a check before the one word in each group to which the definition applies. This may be called a test of care in the use of words. Its main disadvantage is its brevity, since only twelve definitions are used.

Test 5 consists of an exciting extract from Reade's "The Cloister and the Hearth," in which a number of typographical errors occur. The frequency of the errors is directly proportional to the thrillingness of the material; the first part, which is comparatively slow, contains few errors, whereas the last paragraph, which contains one of the most thrilling climaxes in literature, is full of typographical errors. Besides its intended function as a test of one phase of language ability, the test offers some analogy to being assigned to report a prize fight: the excitement of the occasion may cause the reporter to forget his mission.

Test 6 consists of four lists of words, as follows:

I	II	III	IV
thief	to give	assemblage	to say
pickpocket	to award	crowd	to speak
bandit	to donate	audience	to utter

On blank lines below each list the subject enters more words with meanings similar to the three in the list. One great difficulty with this test consists in the scoring. There are a great many words just on the borderline between similar and dissimilar which cannot be scored either way without injustice. Intelligent students probably tend to be more discriminating in the choice of the words which they write down. In scoring, most of the borderline words were considered correct.

Test 7 is constructed on the method suggested by Stern for establishing the validity of testimony of a witness. The

subjects are asked fifty questions covering the details of a picture which they were allowed to look at for one minute. The picture shows a collision between an automobile and a street-car, and is full of detail. The questions cover such items as the license number of the automobile, the amount of damage to the automobile, the number and route of the street-car, the number of the motorman and the policeman, the time of day (shown by a jeweler's clock), the month and day of month (shown by a sale sign over a store), and year (license pad of the automobile). The function tested seems to be related to the reporter's task.

Test 8 is similar in form to Alpha Test 8, but the items are of the kind that a reporter might be expected to know. The following are illustrative:

GOTHAM is a nickname of CHICAGO, NEW YORK, PHILADELPHIA, BOSTON.

IRVIN COBB is a BASEBALL PLAYER, CARTOONIST, AUTHOR, ARCHITECT.

The present POPE is called BENEDICT, PIUS, LEO, INNOCENT.

The United States declared war on Germany in 1915, 1916, 1917, 1918.

MERGENTHALER perfected the LINOTYPE, COMPTOMETER, TELEGRAPH, MIMEOGRAPH.

There are sixty such items in the test.

The student who gets a high score in this test has shown enough interest in his chosen profession to acquire information, through newspapers or periodicals, which will be of use to him later on.

Test 9 is a word-building test, using the letters: A E I O L N R T. All words constructed must have four or more letters. Word-building tests have been classed as measures of verbal imagination.

The total score was computed simply by adding the scores in the individual tests. A correlation coefficient of .97 was found between this method and that of adding ranks in the individual tests. No attempt will be made to weight the tests until they have been finally evaluated.

## 2. The Subjects

This series of tests was given to nine groups of subjects:

- A. 29 Juniors, School of Journalism, University of Washington.
- B. 16 Seniors, School of Journalism, University of Washington.
- C. 47 Freshmen School of Journalism, University of Oregon.
- D. 24 Sophomores, School of Journalism, University of Oregon.
- E. 18 Juniors, School of Journalism, University of Oregon.
- F. 7 Seniors, School of Journalism, University of Oregon.
- G. 85 First and Second Year ~~Non-Journalism~~ Students, ~~Only~~ Washington.

H. 35 Third and Fourth Year Non-Journalism Students, University of Washington.

J. 10 Reporters.

All of these groups except J were tested during January and February, 1920. The first six groups include most of the students in the classes from which they were drawn. At the University of Washington the School of Journalism is open only to upper-classmen. It was impossible to test the Pre-Journalism students in the College of Liberal Arts. Groups G and H were composed of students enrolled in a class in Applied Psychology. They were for the most part in the Colleges of Liberal Arts, Education, and Science. The reporters, who were all on the staff of a metropolitan newspaper, were given the tests under conditions which almost warrant discarding their scores. The men varied greatly in age, education, and experience.

### 3. The Results

While final evaluation of the tests as indicators of success in reporting must be delayed a number of years, a study of the scores may throw some light on the possible future value of the tests.

Some tentative statements about the tests are suggested by the comparison of the median scores for the various groups in each test. Table I gives this information.

TABLE I

MEDIAN SCORE FOR EACH GROUP IN EACH TEST. A SHORTER TIME LIMIT WAS ALLOWED IN GIVING TEST 9 TO GROUPS G AND H. TESTS 3, 6, AND 9 WERE NOT GIVEN TO GROUP J. THE FIGURES UNDER RANGE ARE THE LOWEST AND HIGHEST SCORES ACTUALLY MADE

Test	Group										Range
	A	B	C	D	E	F	G	H	J		
1.....	31.0	37.5	26.6	31.5	21.5	31.0	31.6	31.8	35.5		0-50
2.....	41.7	45.0	33.5	34.5	36.0	39.5	32.3	34.5	31.0		6-61
3.....	16.3	16.5	13.6	12.0	13.0	14.0	10.7	9.0	...		0-31
4.....	7.5	7.9	6.4	6.2	7.0	6.8	6.3	5.9	7.0		1-11
5.....	20.5	32.0	19.0	20.8	18.0	18.0	18.0	15.7	27.5		0-59
6.....	13.7	14.6	13.7	14.3	14.6	14.0	13.3	12.6	...		3-30
7.....	15.0	17.0	12.5	13.4	11.8	11.0	13.0	11.0	13.5		0-34
8.....	21.5	31.5	18.0	19.3	17.5	21.0	18.3	18.3	31.5		3-53
9.....	8.7	9.3	7.3	9.2	8.3	6.5	3.9	3.4	...		0-18
Total Score..	172.5	201.5	149.0	162.5	161.0	152.5	...	...	...		32-293

The University of Washington journalism seniors excel the juniors in all of the tests, but for the University of Oregon students Test 2 is the only one which shows a relationship between median score and year in college.

The University of Washington students of journalism as a whole are distinctly superior to the non-journalism students

in all of the tests. Non-journalism upper-classmen and lower-classmen do about equally well in the tests. In Tests 1, 2, 5, and 8 the reporters do better on the average than the journalism students. Some of the tests seem to distinguish between those who have journalistic inclinations or ability and those who have not.

The correlation coefficients given below were obtained by Pearson's method of correlation of grades, the formula being

$$r = 2 \cos \frac{\pi}{3} (1 - R) - 1, \text{ in which } R = 1 - \frac{6 \sum g}{N^2 - 1}$$

Each test was correlated with Alpha (54 cases), with the following results:

Test	
1.....	.36
2.....	.45
3.....	.58
4.....	.48
5.....	.69
6.....	.30
7.....	.41
8.....	.65
9.....	.43
Total Score.....	.56

These figures are low enough to allow for differences of an appreciable amount if both series of tests are correlated with an objective criterion.

Test 8 in the present series and Alpha Test 8 show a correlation of .62. Tests 3 and 7 correlate .40.

The instructors in journalism were asked to rate their students on the journalistic aptitude which they displayed. These faculty estimates may not be correct, but a student leaving college with the recommendation of the faculty will have his classmates at a disadvantage in obtaining his first position. On the other hand, many students who are considered poor material by their instructors may be given a start by friends or relatives and prove very successful. It remains to be seen what predictive value the faculty ratings have, and whether or not the tests predict the students' future success more accurately.

The University of Washington students were rated by three faculty members and the University of Oregon students by two. With some unimportant variations in wording, the following simple scale was used:

"Rate each of the students who took the journalism tests as follows:

"1, if you think that the student has an aptitude for journalism and if you think that he will succeed in newspaper work.

"2, if you are doubtful about the student's aptitude or do not know him well enough yet to care to pass judgment on him.

"3, if you think that the student has no aptitude for journalism and will not make good in newspaper work."

Each student's ratings were averaged to obtain his final rating. Correlations with test scores are given in Table II.

TABLE II

CORRELATIONS OF FACULTY RATINGS WITH TEST SCORES. ALL FIGURES ARE POSITIVE UNLESS INDICATED NEGATIVE. UNDER AVERAGE ARE GIVEN THE MEANS OF THE SIX COEFFICIENTS FOR EACH TEST

Test	Group						Average
	A	B	C	D	E	F	
1.....	-.20	.34	.29	-.04	.47	-.42	.07
2.....	.13	.45	.31	-.35	.32	.32	.20
3.....	.18	.42	.29	.40	.63	.42	.39
4.....	.43	-.30	.36	.00	.11	.50	.18
5.....	.26	-.07	.28	.35	-.04	.73	.25
6.....	.39	.67	-.18	-.11	.04	.52	.22
7.....	-.08	.54	.26	.24	-.07	.00	.15
8.....	.13	.54	.61	.46	.38	.12	.37
9.....	.13	.62	.09	.57	-.06	.21	.26
Total Score.	.23	.75	.28	.19	.26	.58	.38

These figures show that in general the more advanced the students the higher is the correlation between test scores and ratings. The amount of correlation also varies inversely as the size of the group tested. Supposedly instructors are better able to judge the ability of their students after knowing them longer and meeting them in smaller groups. Selection, however, places a practical limit on the increase in correlation, at least with the present rating scale. If the courses in journalism were extended beyond the customary four years we should eventually find every student rated 1.

Passing to the correlation figures for the individual tests, we find that Test 6 is the only one whose figures show a consistent increase with year in college. Test 8 shows an increase for the Washington students and a decrease for the Oregon students. For Group B Tests 4 and 5 correlate negatively, whereas for Group F they show a high positive correlation. For each of the tests there is one coefficient of  $+ .45$  or more, and for seven of them there are one or more negative coefficients. Tests 3 and 8 have no negative coefficients. We may conclude that as far as this table is concerned, Tests 3, 6, and 8 are the most efficient.



The scores were available for Groups A and B in an intelligence test devised for the purpose of picking officer-material from among naval recruits at the University of Washington during the late war. The test consisted of six parts: cancellation, opposites, memory span, true and false statements, parts of fifty, and perception of number.<sup>3</sup> The correlations between this test and faculty ratings on journalistic aptitude were .14 for Group A and .37 for Group B, in both cases considerably lower than the correlations obtained with the journalism tests.

No Alpha grades were available for any of the journalism students, but many of the non-journalism students had been given the test. The correlation figures on page 53 were obtained on non-journalism students. We may find out indirectly what common factors there are between the ratings on journalistic aptitude and Alpha scores by correlating the coefficients of each test with Alpha given above (page 53) with each of the seven columns of coefficients in Table II.<sup>4</sup> The following are the correlations obtained:

Group	A	B	C	D	E	F	Average
r	.41	-.38	.57	.53	.18	.61	.45

If these figures are taken as probable indices of the amount of correlation between ratings and Alpha, we see immediately that with the exception of Group F the instructors tend to rate their students at first according to their general intelligence and as they know them better they rate them on another ability which is closely related to those measured by the journalism tests.

#### 4. Summary

1. For some unknown reason the tentative criteria show more positive results with the University of Washington students of journalism than with those of the University of Oregon.

<sup>3</sup> The "parts of fifty" test consisted of twenty rows of eight figures each. Two figures in each line added together equaled fifty. The subject was required to cross out these two figures in each line. The "perception of number" test consisted of sixty rows of circles. The subject was required to write on a blank line after each row the number of circles in the row. The circles were at even distances apart, so that it was not necessary to count out the number of circles in every line. In case a line contained more circles than the preceding one, it was merely necessary to add the number of circles extending beyond the preceding line.

<sup>4</sup> For a theoretical discussion of this method, the reader is referred to Hart and Spearman: "General Ability, Its Existence and Nature," *Brit. Jour. of Psychol.*, Vol. V, 1912.

2. Tests 1, 2, 5, and 8 tend to distinguish between reporters, students of journalism, and non-journalism students.

3. In general, faculty ratings on the aptitude of their students agree closer with test results when the instructors meet their students in small groups than when they meet them in large classes.

4. In general, the longer the instructors know the students, the closer do their ratings agree with the journalism test scores.

5. The scores in Tests 3 and 6 show the most consistent agreement with faculty ratings.

6. The results lend support to the analogy method of devising tests.

The final evaluation of the tests must be delayed until the students have been out of college for five or ten years. In grading them on their success a rating scale will be used which will contain such items as the importance of the newspaper on which the man is employed, importance of his position, his salary, etc.

A revised edition of the test series was printed during March, 1920, and was given to students of journalism at the Universities of Minnesota, Illinois, Indiana, Kansas, and Iowa, and Ohio State University. Copies of the revised series may be obtained by addressing the writer at Carnegie Institute of Technology, Pittsburgh.

## TESTS IN INDUSTRY\*

By MORRIS S. VITELES, A. M., University of Pennsylvania

The vocational psychologist concerns himself with finding the right man for the right job. There are two avenues of approach to this: that of choosing a worker to fill a particular job and that of choosing the job which is most suitable for the particular worker. Industry is interested in getting the worker for the job and concerns itself little with the latter and more difficult problem which confronts the vocational counsellor. Both problems, however, involve an analysis of jobs from the point of view of the specific mental abilities which the worker must possess in order to attain proficiency in any one of them and the development of scientifically accurate measures of such ability. "Tests are the devices by which mental abilities can be measured." Industry has lagged far behind education in adopting a quantitative standard for the measurement of the ability of individuals. It has selected the worker haphazardly and the function of selection has often been assigned to one of the menials in the executive hierarchy of industry. Link, in his *Employment Psychology*, cites an amusing example of this haphazard method of selecting workers. "One large manufacturing concern," he writes, "has appointed as monarch of the 'hiring-on-window' a man who had the misfortune to lose a leg in the company's employ. As a consequence of this loss he was given his present life job which he performs to the queen's taste. He was induced to describe his methods and they were something as follows. 'On Monday I turns down all the men with white collars, on Tuesdays all with blue eyes, Wednesday all with black eyes. Red-headed men I never hires, and there do be days when I have a grouch and hires every tenth man.'" This is perhaps an exaggeration, but it illustrates the haphazard method of selection which industry still largely employs. Numerous examples can be cited which illustrate the important part played by the personality and tastes of the employment agent in selection. The head of a billing department of a

---

\* An address delivered at the Conference on "Tests in Relation to Vocational Guidance" at the annual meeting of the National Vocational Guidance Assn., Atlantic City, Feb. 26, 1921.

department store for which I was preparing tests was praising his force and, took me on a personally conducted tour of his department to point out his star workers. "Now this girl," he said, pointing to one, "is going to make a cracker-jack worker. I just hired her this week, but I know it. I know that every time I hire a girl with that kind of a face, round, serious looking, smart, she makes a fine biller. Now I must have been asleep when I hired that one. She won't make good," and he indicated a petite, dark-haired girl whose features, in my opinion, radiated intelligence and ability. The girls whom he had indicated as the possessors of pleasing, intelligent faces to me seemed dull and lifeless. I am free to admit that it was altogether a question of personal tastes, but it was such personal likes which dominated the choice of workers for this department, and for other departments in the store—and resulted in the choice of personnel lacking uniformity and with greatly varying competency for the different jobs in the store.

It is by reason of this absence of uniformity in selection that tests are replacing in industry the haphazard method of selection in which the only measure of the applicant's fitness for the job is the judgment of the foreman, or even the better trained judgment of the employment manager on the basis of mere observation. Sometimes those who make the selection are using such additional criteria as the contour of the applicant's face, whether it is convex or concave, the length of the fingers, the texture of the skin, the shape of the head and any number of other unreliable standards which pseudo-scientists are recommending to industrial executives. The readiness of industry to employ these methods is an indication of the need which is felt for some more definite criterion of selection than that of mere observation.

Two kinds of tests are being used by industry to provide these more definite standards of selection;—trade tests and psychological tests. The former are being used to determine whether the applicants for a job in which experience is required actually possesses such experience. They are measures of trade experience involving tasks which only an experienced worker in the trade can perform. The carpenter is given a block of wood, a hammer and a chisel and instructed to make a certain kind of a joint. The inside electrician is given material and is instructed to put up some wiring. Such tests were used extensively in the army during the late war to pick the skilled workers to build cantonments, do electrical wiring, run auto trucks, etc. The psychological tests, on the

other hand, are such as measure the ability of the inexperienced applicant for a position to attain proficiency in the job for which he is applying. The dividing line between the two types of tests is not a clear cut one at times, but the fundamental difference is an essential one. I shall confine myself, in this paper, to a discussion of the psychological tests, with particular reference to their use in selection, and not for promotion, transfers, the regulation of wages and any number of other uses to which they are put in industry.

The psychological tests, may be divided into two classes, tests of general intelligence and tests for specific abilities. The general intelligence test, such as the Army Alpha, the Binet Simon test, the Otis test, and many of the so-called intelligence tests which are being used to grade school children are really measures of the performance level of the individual. The college professor stands higher on the performance level than the skilled tradesman, and the skilled worker on a higher performance level than the laborer. The higher the performance level of an individual the higher the score he will make on a test for general intelligence, which may be defined as a combination of innate specific abilities, educational and social opportunities. Such tests have a very limited use in industry. They are used in the first place, to shut out from employment the feeble-minded, those who are altogether unfit for any job in the plant by reason of deficiency in mentality. They are also used to select workers for jobs in which success depends, to a very great extent, upon a high level of general intelligence, namely the executive positions. For selection for the great mass of skilled and semi-skilled jobs, for office jobs and simple clerical jobs, the general intelligence test cannot be used. The carpenter and the toolmaker, for example, must stand on approximately the same level with reference to general intelligence, but very different specific abilities are required for these trades.

In an analysis of the tests for general intelligence given in the army it was found that laborers and farmers averaged C—in general intelligence; bricklayers, cooks, bakers, general machinists, shop carpenters and workers in a number of other trades averaged C; clerks, photographers, R. R. clerks, etc., averaged C+. If trade competency correlated with general intelligence it could be predicted that everybody who ranked C in the army alpha would make an equally good bricklayer, cook, baker or general machinist. It is a matter of common experience that many bricklayers could not be trusted to prepare a meal successfully even after much training, and that

many first class chefs could be taught bricklaying for many years and would never succeed in aligning properly a row of bricks. Trade competency is related to specific abilities which the individual possesses. Specific abilities are those innate traits which enable an individual to perform some specialized task. In the schools we find good pupils who are poor in arithmetic, good arithmeticians who do not learn to spell well, and individuals who are poor in spelling, arithmetic, reading and all other school subjects who do remarkably well in the classes in carpentry and clay modelling. A special set of specific abilities go into each of these, and differences in the specific abilities which individuals possess account for differences in arithmetic, spelling, reading competencies and so forth. In the same manner competency for different jobs in industry depends upon specific abilities. A motorman on a trolley car must possess a high degree of distribution of attention, alertness, planfulness and the ability to react quickly with a learned response to external stimuli. The conductor, on the same car, needs distribution of attention, alertness, planfulness in a more limited degree, but the last complex ability is not needed. The conductor on certain roads needs a certain amount of arithmetical competency, the motorman need possess none of this. The general intelligence needed for these two jobs is about the same, but there are many men who make good conductors who would be dangerous characters as motormen by reason of deficiency in the ability to respond with certainty to the stimuli with certain learned responses. A girl who wraps locks must concentrate her attention on one particular object to the exclusion of all others; a girl who inspects locks at the same table must distribute her attention over a number of different elements of the locks. Concentration and distribution of attention are two distinct specific abilities. It was found in one plant that girls changed from one job to the other failed to make good and a psychological analysis of the job explains this failure on the basis of the different specific abilities involved in these jobs. The tests which are being used successfully in industry are those which measure the specific abilities of individuals for particular jobs, and not tests for general intelligence.

The process of developing tests for industry is a long one. It involves, first of all, an analysis of the job from the point of view of the abilities involved in its performance. Such an analysis can only be made by a psychologist. In some cases it can be made by merely observing the worker on the job, but in most cases it is actually necessary for the psychologist

to become a worker, to don the overalls of the machine hand, the gloves of the motorman, the green shade of the clerk if he desires to make an adequate analysis—to understand<sup>1</sup> “the mental action for which he is trying to find tests as measures.” The next step is to find or devise tests which, in his opinion, involves in their performance the same specific abilities which are involved in doing the job. For this purpose the psychologist can use, sometimes with only slight modifications many of the tests which are in common use, but it is often necessary for him to construct an entirely new piece of apparatus. Analyzing the job and preparing the test are only preliminary steps in developing the tests for particular jobs. There are two additional steps which must be carried out. In the first place the fact that the test is actually a measure of the abilities which the job requires must be empirically demonstrated. In other words, the tests must be tested. This is generally done by submitting workers who are already employed on the job to the test and finding out whether the more proficient workers actually pass the test with a higher score than the less proficient workers. Sometimes, a number of tests which seem to involve the same ability are given, and these selected which show a high correlation with proficiency in the job. Another and longer but better method is to test all applicants for the job with the tests which have been selected by the first method, to make a record of their scores on the test, and to follow them up after they are employed at specific intervals to determine whether the individuals who made high scores become proficient workers and vice versa. (In this connection, of course, it must be recognized that there are other factors than inability for which an individual is found unsatisfactory in industry, and that the test measures only this, the most important factor in satisfaction.) The other step is to provide quantitative standards for selection. The standards must be in terms of units or of times, and the exact point which employment is advisable must be indicated. Such points, critical levels they are called, can only be established empirically, and as more applicants are tested the standards so established are revised and refined.

It is necessary to recognize that tests which are being used in industry overlook an important feature—temperamental differences between individuals. Two men may have the same specific abilities; both may have equal competency for a particular job, but one may be a willing worker, the other may be an unwilling worker. One may be tempera-

---

<sup>1</sup>Link. *Employment Psychology*.

mentally fitted for the job, the other may not. The first may be fond of indoor work, the second of outdoor work. The vocational psychologist has hardly attempted to measure these temperamental differences, largely because little of concrete value has been accomplished in the analysis of mental qualities which are called temperamental. Both for purposes of selection in industry, as well as in vocational guidance, definite devices for measuring temperamental differences will have to be developed before selection can be extremely satisfactory.<sup>2</sup> "It must certainly seem," says Watts in an article in the *British Journal of Psychology*, "that the vocational psychologist will need to take into consideration what has so far been overlooked, the instinctive make-up of the subjects of his experiments, and that in choosing workers for particular tasks he must make sure that their instinctive type is the right one. It would be folly to place the man with strong gregarious tendencies in lighthouse or wireless work, or advise him to go into farming in a new colony, just as it would be unwise to encourage a boy devoid of the acquisitive instinct to set up in business. And it will be agreed that a soldier or a reformer without a considerable spice of pugnacity would be a failure, while it is equally beyond dispute that the statesman and the priest, the doctor and the nurse, the teacher and the shepherd would soon tire of their work if they were not endowed strongly with the tender impulse." The next few years will show whether we shall be able to devise tests for the diagnosis of the relative strength of man's various instinctive qualities. It would be possible to quarrel with Watt's analysis of instinctive qualities, but the essential problem is well stated. In addition to abilities there are motives and purposes in the worker's make-up. There are, as has been popularly stated, "'doers who will' and 'doers who won't' and the 'willer' and the 'wonder' present another problem for quantitative treatment with tests." That judgment of temperament made by observation of facial characteristics, color of hair and other physical features, which is recommended and used by popular vocational counsellors, is unreliable is very apparent to the scientist.

One factor in vocational guidance should be the child's competency for the occupation which he seeks or is being advised to enter. The judgment of the counsellor upon the presence of the competency is as unreliable as the judgment of fitness for a job made by the employment agent in indus-

---

<sup>2</sup>F. Watts, "The Outlook for Vocational Psychology". *British Jour. of Psychol.*, Jan., 1921.



try. The presence or the absence of such competency can be determined by the use of the same tests which are being used by industry in the selection of its workers. Associations interested in vocational guidance should co-operate with industry and obtain from industrial plants the tests which they are using. Such an association should also undertake an analysis of jobs to determine what constitutes competency for particular jobs and the development of measures of such competency, in the same manner as industry is doing. We have passed the time in vocational guidance when jobs can be defined in terms of such abstract qualities as good, hard, clean, desirable, etc. We have reached the point where an accurate definition and reduction to measurable terms of job competency is absolutely essential.

In the use of tests for vocational guidance it must be recognized, however, that competency for a job, as measured by a test, is only one factor in guidance. It is combined with the child's health, his desires, his temperament, and other conditions to determine final choice of an occupation. The test, however,—the same test that industry uses in the selection of workers for its jobs can be used in the place of the harsher method of trial and failure to weed out those who are definitely incompetent to attain proficiency in vocations which otherwise seem suitable and to point the way to those whose desires are not well defined.

# THE PROBLEM OF THE UNSELECTED GROUP IN THE STANDARDIZATION OF TESTS

By S. L. PRESSEY

Studies from the Psychological Laboratory, Indiana University

## I. *The Essential Impossibility of Obtaining an Unselected Group in Work with School Children*

The difficulty of obtaining satisfactory norms or standards for adults for any type of test—of general intelligence, vocational aptitude, or what not—has come to be generally recognized. It is realized that different groups of adults vary astoundingly, in the first place, in the extent to which various selective influences have operated to bring them together in this particular sample. The subtle extent to which environmental influences may also come to have an effect upon a person's performance on such tests is also coming to be appreciated. But it seems to be supposed, by some workers in the field, that school children make material for thorough standardization which may be considered free from the influence of such special factors. If there is one single fact which has surprised the writer, from the study of school surveys made during the past three years, it is the fact that such influences are extremely important and pervasive in the public schools. And he has come to the conclusion that standard groups must be very carefully defined if the standards are to mean what they should for comparative purposes. The purpose of the present paper is to present very briefly a few bits of data illustrative of the way in which "sixth grade norms" from one school system may mean a very different thing from "sixth grade norms" from another, nine year old norms in one school may mean a very different selection of nine-year-olds from the group that would have been gathered under similar circumstances in another school—and so on. The data are, to the writer at least, extremely interesting and involve, he believes, matters of extreme importance.

## II. *Age-Grade Distributions and Mental Tests*

As illustrative of two types of school system the two following age-grade distributions are presented (see table I).

The results are the age-grade tabulations obtained from test blanks used in a survey of these two school systems made by the writer and his wife. In this work, the "Primer" scale was given to the children in the first three grades, and the "Cross-out" scale to children in the grades from the fourth through high school.<sup>1</sup> Age tabulations from these blanks thus result in a complete age-grade census of the entire school system at the time when the survey was made. The testing was in each case all done within three days and thus gives a very satisfactory cross-section statement of the situation in these school systems at the time of the examination. System C was examined the last of February, System F about three weeks later, in the middle of March. These age-grade tables are full of interesting facts, only a few of which can be mentioned here.

TABLE I  
AGE-GRADE DISTRIBUTIONS

City C.															
	1	2	3	4	5	6	7	8	9	10	11	12	Total		
5	2	...	...	...	...	...	...	...	...	...	...	...	2	5	
6	60	2	...	...	...	...	...	...	...	...	...	...	62	6	
7	60	40	2	...	...	...	...	...	...	...	...	...	102	7	
8	34	57	37	2	...	...	...	...	...	...	...	...	130	8	
9	9	30	60	35	1	...	...	...	...	...	...	...	135	9	
10	1	22	19	58	29	...	...	...	...	...	...	...	129	10	
11	...	6	14	41	50	21	1	...	...	...	...	...	133	11	
12	...	3	8	29	39	38	25	1	2	...	...	...	145	12	
13	...	...	3	9	22	28	29	24	8	...	...	...	123	13	
14	...	...	1	4	2	18	28	25	38	3	...	...	119	14	
15	...	...	...	...	5	5	4	13	27	15	2	...	71	15	
16	...	...	...	...	...	1	5	25	25	14	3	...	73	16	
17	...	...	...	...	...	...	...	1	12	16	27	...	56	17	
18	...	...	...	...	...	1	...	1	1	9	16	...	28	18	
19	...	...	...	...	...	...	...	...	...	...	8	...	8	19	
20-21	...	...	...	...	...	...	...	...	...	...	1	3	4	20-21	
Total	166	160	144	178	148	110	89	68	102	56	42	57	1320		
City F.															
	1	2	3	4	5	6	7	8	9	10	11	12	Total		
5	...	...	...	...	...	...	...	...	...	...	...	...	...	5	
6	105	3	...	...	...	...	...	...	...	...	...	...	108	6	
7	52	65	...	...	...	...	...	...	...	...	...	...	117	7	
8	12	69	74	3	...	...	...	...	...	...	...	...	158	8	
9	...	11	58	71	2	...	...	...	...	...	...	...	142	9	
10	...	4	22	62	58	3	...	...	...	...	...	...	149	10	
11	...	...	...	26	67	61	...	...	...	...	...	...	154	11	
12	...	...	...	10	29	66	53	...	...	...	...	...	158	12	
13	...	...	1	5	13	40	47	50	...	...	...	...	156	13	
14	...	...	1	1	7	16	41	53	36	2	...	...	157	14	
15	...	...	...	1	1	2	14	22	35	28	2	...	105	15	
16	...	...	...	...	...	...	...	12	8	22	24	1	67	16	
17	...	...	...	...	...	...	...	1	1	4	24	18	48	17	
18	...	...	...	...	...	...	...	...	1	2	4	11	18	18	
19	...	...	...	...	...	...	...	...	...	...	3	3	6	19	
Total	169	152	156	179	177	188	155	138	81	58	57	33	1543		

<sup>1</sup> See Cross-out Tests, *Jour. of App. Psychol.*, June, 1919; and A Brief Group Scale for Use in the First Three Grades, *Jour. of Educ. Psychol.*, Sept., 1919.

First of all, the question is as to the comparability of grade norms obtained from these two systems; the grade norms are naturally the first to be worked out, since the data comes in grouped by grade and not by age. The grade distributions given above look roughly similar; but comparison of the median ages for each grade brings out, at once, differences of decided importance. These median ages run as follows:

	1	2	3	4	5	6	7	8	9	10	11	12
City C..	7.35	8.68	9.55	10.90	11.88	12.90	13.64	14.36	15.11	16.40	17.31	17.95
City F..	6.80	8.12	9.06	10.25	11.42	12.45	13.52	14.36	15.13	15.97	17.10	17.86
Diff....	.55	.56	.49	.65	.46	.45	.12	.00	-.02	.43	.21	.09

These differences expressed in hundredths of a year perhaps do not look very surprising. But when it is realized that the children in the first grade of City C are over six months older than the children in the same grade of City F and that the children in the fourth grade of City F are eight months younger than the City C fourth graders, then one realizes that grade norms from City C and City F are *not* comparable. Just what factors bring about these results it is hard to say. It may be that the children entered school later in the second city. Or they may be held back more in their school progress. But whatever the causes may be, the result is unmistakable; grade for grade the children of City F are about five months older than the children in City C. If one is working with tests of mental endowment the situation is bad enough; allowance must be made throughout for this difference in maturity. If the work is with tests in the school subjects, the situation is almost hopelessly complicated. If the children in the second city be found to test higher, this result might be due either to greater maturity, or greater time in school.

Any differences might in fact be due to selective factors connected with these facts. Close study of the two distributions will show that in the first city the dull children tend to be held in the first three grades. But in the second city the incidence of retardation is between the fourth and fifth grade. And selection for the upper grades is much more severe in the first city. There are practically the same number of children in the first grade of City C as of City F (166 and 169 respectively). But only 41 per cent of this number are in the eighth grade in the first city compared with 62 per cent in the second city. It might almost be said that the eighth grade children in the first city were twice as highly selected.

When norms are said to be for the eighth grade, then, in order to have the nature of the group classified, the median

age of the grade should be known, the amount of elimination and retardation—in fact, the total age-grade distribution for the school. In one way or another the total policy of the school system has managed to express itself to an extraordinary extent in merely the crude fact of the age-grade relationship.

So much for grade norms; in dealing with age norms the situation is much more complex. The writer would like to consider particularly the problem of standardization as it appears in work with any one of the group scales for work with the upper grades now available. Such scales involve the ability to read, and so are not very satisfactory in work below the fourth grade. The question is then as to whether satisfactory norms can be obtained by age if children only above the third grade are examined. It is also interesting to consider the influence upon the norms of the admission of high school. In the distributions already referred to a break is made between the third and fourth grade and between the eighth and ninth in order to set off the grade school group most naturally dealt with in work with such scales. A glance at the distribution for City C will reveal the extraordinary fact that if only the children from the fourth through the eighth grade had been examined, it would have been impossible to obtain complete age distributions for *any* single age. Only for ages 12 and 13 would a complete distribution be approximated. Even if we include the third grade we do not manage to include all the cases of any one age. And to present a series of age norms based on data thus obtained approaches rank absurdity. Such norms are warped and twisted out of all shape by these selective influences in the school system.

In City F there is a somewhat better situation. By including the third grade all the 11, 12, and 13 year old children would have been examined. But it is obvious that age norms obtained from this city would not mean at all the same thing as age norms thus obtained from City C. And there is a further factor here of great interest. In City F practically all the fourteen year old children are still in school; there are 158 12 year olds in the system and 157 14 year olds. But in City C there are 145 12 year olds to 114 14 year olds. This elimination had begun to show, in fact, even at thirteen; there are about twenty less 13 year olds in City C than we would expect. Comparable age norms from the two cities would seem, if a scale simply for the upper grades is used, an utter impossibility. It is also interesting to notice that

age at entering school apparently varies, and that the lower ends of the distributions are also different. Thus the first city shows only half the expected number of six year olds in the school, where the second city shows 75 per cent. The situation seems hopelessly special, and the obtaining of a general standard next to impossible.

It is especially interesting to consider in this connection that it is altogether impossible, from work in the schools, to obtain any unselected adolescent norms. It is perfectly obvious that the children 15, 16 and 17 years old who are in school are simply a fraction of the total number of children of these ages. How much they have been selected it is hard to say. The comparative numbers for these ages for the two cities mentioned are by no means so different as are often found. Thus the number of the 16 year olds still in school is 51 per cent of the 12 year olds in City C. But in a third system examined at almost the same time, there were 73 per cent of the expected number of 16 year olds in the public schools. This lack of norms for adolescence is of decided importance. The army psychological work did not, of course, test this group. And the fundamental fact is that we simply don't know anything about what performance we should expect of our delinquent girls, juvenile offenders, or the large and perplexing group of adolescents just going to work.<sup>2</sup>

### III. *The Age-Grade Description of Standard Groups*

In the light of such data it at once becomes obvious that in presenting norms for any type of test some statement of both age and grade grouping should be included. There are,

---

<sup>2</sup> It need hardly be said that the results from a single school must—if results from an entire school system are unsatisfactory—be largely out of the question as a basis for standardization. A town or city is more or less of a sociological unit. In such a total community a proportionate representation of all classes may be expected. But a single school may draw largely from a neighborhood made up almost altogether from one class of people; one school may contain almost altogether the children of day laborers, another be in an excellent residence district. As an example, selected almost at random, schools X and Y in a small Indiana city may be instanced. The median ages for the third, fourth, and fifth grades run 9.07, 9.55, and 11.00 for the first school and 8.06, 9.28, and 10.55 for the second. The first school has 32 in the third grade as compared with only 18 in the fifth; the second has 62 pupils in the third and 87 in the fifth. On the basis of such data both age and grade comparisons seem, at first thought at least, utterly impossible if anything more than school selection is being investigated. (See also "A Comparison of Two Cities and Their School Systems by Means of a Group Scale of Intelligence" *Educational Administration and Supervision*. Vol. 5. 1919. pp. 53-62.)

of course, a number of ways in which this may be done. The thoroughly scientific way would be to present the age-grade distribution; but a summary statement of the age-grade situation would still be needed. It is becoming a somewhat common practice in work with educational tests to give at least the time of the year when the tests were given. There is some effort now to obtain the age norms for these tests in the various school subjects. The writer cannot see how, in view of the various facts which have just been presented, this can be done very satisfactorily. Tests in the various school subjects are rarely applicable over a wide range of ages or grades; under such circumstances the obtaining of age norms which are reasonably free from the influence of grade selection would seem altogether impossible. The writer would like to make the following suggestions.

(1) In work with tests of general ability grade norms should be invariably accompanied by the median age, for each grade, of the standard group from which the norms are obtained. Time of examining in the school year is relatively unimportant. A statement of the median age will permit allowance for examining at different times of year, and allowances also for differences in amount of retardation and acceleration. It is probably the best single statement of the age-grade situation.<sup>3</sup> In dealing with tests in the school subjects or influenced by the school training, both median age and time of examination should be given.

(2) In developing age norms either tests should be given to the entire school system, or a census of the entire system, at time of examination, should be taken. The number of cases at each age outside of the group examined should be determined and allowed for in the norms so far as possible; such cases may be counted in at the top and bottom of the distribution if the number of such cases is small enough to make such procedure safe. In determining whether the entire population at a given age is included, age 12 is probably the best reference point. Twelve year olds are very rarely found in

---

<sup>3</sup> Statement in terms of per cent of retardation is, after all, a very clumsy statistical method. It requires in the first place reference to a standard time of the school year; September 1st is usually taken. And it, of course, requires that the census be taken at some other standard time of the year. Limits and standards of retardation and acceleration differ, normality may be granted only to children within a range varying from six months to three years. After all the central tendency is the important matter, but it is dealt with in terms of number at one extreme. Why not simply median age and either inter-quartile range, ten and ninety percentiles, or something like that? The whole situation might be clarified by some such method.

high school, are usually not yet eliminated; and it may be considered reasonably sure that all the twelve year olds have already entered school, and that most of them have progressed far enough to be dealt with by the usual test. If possible, it is highly desirable that a test be given to the first three grades. Otherwise a perspective on the entire situation cannot be gained.

(3) In developing norms for either age or grade, recognition must be taken of the fact that different schools and neighborhoods differ markedly in age-grade distributions, in the ability of the children, and in a variety of other ways. Differences from school to school are likely to be much more striking than differences from one community to another, since segregation of different social classes in different districts of each community is frequent, but entire communities contain representatives of all classes. Standardization on the basis of entire small school systems is, therefore, much more satisfactory than standardization from the same number of children in two or three schools of a large city. In the latter case it is impossible to determine the influence of sociological selection upon the norms.

(4) Our knowledge of standard performance is most woefully lacking for the adolescent group from fourteen on. In the present state of knowledge it is probably best not to attempt any standard norms, but simply to develop norms for school children, for working children, and so on. The greatest single need to-day in the field of mental tests is, perhaps, for the adequate determination of the total distribution of ability in adolescence.

### *Summary*

The paper deals with the problem of obtaining unselected, representative, groups of cases for the determination of norms on the various types of tests. It is shown that

1. Grade norms are highly unreliable because of differences in amount of over-ageness, elimination, and age at starting to school, from one school and one school system to another.

2. Unselected age groups are extremely difficult to obtain since (a) children of a given age may be found scattered over as much as eight grades, and (b) the exact nature of this grade distribution varies from school to school and community to community.



3. It is pointed out that no satisfactory norms are available for ages above fourteen and that elimination from school, at these ages, varies markedly from one school to another.

4. It is suggested that (a) in presenting grade norms median age per grade be invariably given, (b) in presenting age norms a complete age-grade census of the entire school or system should be obtained. The need for careful selection of groups for standardization, and the present inadequate knowledge of adolescent performance on mental tests, is also pointed out.

MINOR STUDIES FROM THE PSYCHOLOGICAL LABORATORY OF INDIANA UNIVERSITY

VIII. A PRELIMINARY INVESTIGATION OF GENERAL PROGNOSIS—I. E.,  
"GENERAL INTELLIGENCE"

By S. L. PRESSEY

1. *Problem and Materials.* The present study could have been much better made in connection with some investigation in the field of business psychology; the greater variety and the more specific nature of the abilities required in any business concern, and the definite monetary criteria, would have made the analysis much more valuable. As it is, the materials used have to do with the public schools simply because these data were most available. The method was worked out, in the first place, with the notion that it might prove of value in personnel work. In a previous paper in this JOURNAL<sup>1</sup> data regarding the efficiency of a group scale of intelligence, in prognosticating failure and success in Junior High School, were presented. In the discussion regarding this material it was suggested that extended investigation regarding the specificity of prognosis problems was much needed. That is, for the purpose of foretelling success or failure, in the endeavor to more closely co-ordinate work and ability, it is quite conceivable that different tests and methods may be necessary for long and for short time prognosis, or for prognosis in this or that line of work. It was urged that only by such investigations could a decision be reached regarding the validity of the "theory of general intelligence—that is, the theory that prognosis problems can be generalized." The present brief study is, first of all, a preliminary essay at investigation of this problem; it is, in the second place, an attempt to relate, in a very concrete and direct fashion, test score and practical criteria.

The data consist of (a) the scores made two years ago by the children in a sixth grade Junior High School class, on a group scale of intelligence previously described in this JOURNAL<sup>2</sup> and (b) the marks received by these same children at the end of the seventh grade, a year and a half later. A total of 115 children of this class took the examination and also had the complete school records required by the problem. And the very fundamental question is this: Can success and failure be foretold "in general" (so to speak) by the scale; or, are different tests needed according as one is trying to foretell success in different subjects? As subjects largely contrasted, in content, interest, and type of function which seems to be involved, arithmetic and English have been chosen for study.

2. *Results.* It was desired to relate the tests and the criterion in as practical, and as concrete and intimate a fashion as possible. Under

<sup>1</sup> December, 1919.

<sup>2</sup> Pressey, S. L. and L. W. A Group Point Scale for Measuring General Intelligence, *J. of Appl. Psychol.*, Vol. II, 1918, pp. 250-269.

the circumstances, a series of correlation coefficients could hardly yield the information desired; it is tardily being realized that a correlation surface may be of an infinite variety of shapes, that for purposes of individual diagnosis summary of such tables in a coefficient may be of comparatively little value—that to many scatter-diagrams a correlations formula is, essentially, inapplicable.<sup>3</sup> It was, therefore, decided that relationship over the entire range of the distributions should not be studied, but only the relationship of the tests to failure—that is, relationship about the passing mark (75). The following table shows, in a simple and direct way, the efficiency of the scale

		TEST SCORE											
		45	50	55	60	65	70	75	80	85	90	95	
Failed.....		..	..	..	1	..	..	..	3	3	3	..	
Passed.....		..	..	1	1	..	..	..	4	2	5	6	
Total.....		..	..	1	1	..	..	..	7	5	8	6	

		TEST SCORE—Continued										
		100	105	110	115	120	125	130	135	140	145	Total
Failed.....		3	3	..	2	4	..	..	1	..	..	22
Passed.....		9	5	15	19	4	8	4	6	3	1	93
Total.....		12	8	15	21	8	8	4	7	3	1	115

in marking off those who have failed in two or more subjects. As will be seen, by no means all the failures made poor scores on the scale, and by no means all those who made poor scores failed at this particular time. But in general the failures are to be found at the

<sup>3</sup> See for instance Kornhauser, A. W., Meine, F., and Ruml, Beardsley. Two models showing the interrelation of several significant correlation variables. *Psychological Bulletin*, Vol. XVI., 1919, pp. 230-235. Thurstone, L. L. Tests for College Entrance, *Journal of Educational Psychology*, Vol. X, 1919, pp. 129-42.

With regard to the writer's data it may be said shortly that the distribution of marks was different for each subject; thus the range for arithmetic was very narrow (over half the marks were between 75 and 80), the distributions for English, Spelling, and Geography were bi-modal, Spelling was skewed very markedly to the top and Geography as definitely toward the bottom. Evidently each teacher had her own standards, her own notions as to the proper allotment of marks; the teachers also presumably differed in the care with which they assigned their gradings, and differed also in their ideas as to what, after all, school marks should mean in the way of recognition of ability, reward for effort, hopefulness for a child's future—and so on! But there is one point in the distribution about which a teacher must think seriously, about which she is much less left to her own devices—and that is the passing mark. Most important for the practical purposes of this study—it is the one point with reference to which the school takes practical action. The passing mark—75—is, therefore, taken as the reference point in the statistical work.

lower end of the distribution: in fact, the relationship is perhaps about as close as could be expected, all things considered.

Such a table as the above exhibits, in a very obvious and direct fashion, the relationship between test and school record. But for comparative purposes some way of summarizing such a table in a single index number is evidently necessary. The form of statement finally used was quite as simple and direct as the table from which it was obtained. If (it was argued) the scale had been, in this particular instance, a perfect prognostic instrument then all the failures would have been at the bottom of the distribution of test scores; the scale would have been 100% differential. As a matter of fact only 9 of the 22 failures are among the 22 who score lowest on the tests. That is, 9 out of the 22, or 41%, are "where they ought to be;" the scale is about 40% efficient, for this particular purpose. Such a statement (of "per cent where they ought to be") is easily obtained, is largely independent of the number of failures dealt with, and would seem very well suited for the rough comparisons dealt with in the present paper.<sup>5</sup> And it may be said at once that the remaining data of the paper are presented in terms of these percentages.

The above results will serve to indicate something as to the prognostic efficiency of the scale "in general"—that is, the efficiency in forecasting those who will fail in *any* two subjects. But the writer was interested rather in efficiency along specific lines, in prognosticating failure in particular subjects. The failures in English (19 in all), and in arithmetic (22 in all) were, therefore, tabulated in the same fashion, and similar per cents of efficiency found. It appeared that the scale was 37% efficient so far as English was concerned, and 22% efficient in setting off failures in arithmetic.

However, for adequate details not only failures in each subject, but results on each test, should be considered. The following table shows the per cent of efficiency of each test, in differentiating failures in English, and in Arithmetic.

	TEST									
	1	2	3	4	5	6	7	8	9	10
English.....	5	31	26	11	21	26	16	37	37	26
Arithmetic.....	21	18	29	36	21	32	14	21	29	29

<sup>4</sup> The complexity of the factors influencing a child's school record at a particular time must be constantly kept in mind. Thus of the 19 cases who passed, but who scored below 100 points on the scale (the 25 percentile) 7 had failed of promotion during the previous year, and were progressing now only because they had recently repeated some of their school work. Two of those who failed, but who made a good score on the tests, are definitely known to have failed because of poor health. But, in general, there is a positive relationship between test and school success and failure.

<sup>5</sup> In comparing differential efficiency for different subjects the total distribution of test scores was first tabulated, once for all, from the card records. It was then only necessary, in studying each subject in turn, to retabulate only the scores of those failing. By marking in the failures with red pencil so that they stood out clearly—and by keeping the slide-rule handy—the whole method was extremely easy. Research "sketching" one might call it.

Apparently (if we may trust this table) the tests differ distinctly in their ability to prognosticate success in these two subjects. And the practical question at once arises as to whether, by judicious selection among these tests, a shorter scale, or "differential unit," more efficient than the entire scale, might not be obtained. To make trial of the matter the four tests most differential, for each subject, were selected, on the basis of the above table, and the tabulations from which it was derived.<sup>6</sup> And the scores made on these four tests by each child were added together, and the sums tabulated after exactly the same fashion as were the total scores on the entire examination. Thus the four most differential tests for failure in English were 2, 3, 8, and 9.<sup>7</sup> And Floyd T. made scores on these four tests of 8, 4, 8, and 11 points; his score on the English sub-scale was, then, 31 points. The tabulation showed this sub-scale to have a prognostic efficiency of 47%, as compared with 37% for the entire ten tests of the total examination, in differentiating failure in English. And the arithmetic sub-scale (tests 4, 6, 9, and 10) gave 36% efficiency in differentiating failures in arithmetic, as compared with 22% for the whole scale.

Apparently the tests of the entire scale differ distinctly in their prognostic value. There are, of course, two possible explanations. Either certain tests are efficient for one purpose but not for another; or, certain tests are of relatively slight value for any purpose. As has already appeared, the best tests for dealing with English are by no means the best tests also for differentiating potential failures in mathematics. However, these sub-scales yield further information on this point. The arithmetic sub-scale is 32% differential of failures in English, and the English sub-scale is only 15% differential of failures in arithmetic! That is, there is a reversal; tests may, apparently, be excellent for one purpose and poor for another. In an effort to emphasize this feature a sub-scale demonstrating this was sought. Analysis by test showed tests 2 and 8 to be differential for English while poor for arithmetic, and tests 1 and 4 to be efficient for arithmetic but poor for English. A sub-scale made by combining tests 1 and 4 gave 11% differential efficiency for English and 29%

<sup>6</sup> Where tests were tied for fourth place, as is the case with tests 3, 6 and 10 for English, and tests 9 and 10 for arithmetic, choice was settled by inspection of the original tabulations.

<sup>7</sup> The nature of the tests is unimportant for the present discussion; the point is only that tests appear to differ in their efficiency for different purposes—that the prognostic problems dealt with are, to a certain extent at least, specific problems. For a detailed description of the tests the reader is referred to the original article presenting the scale (see note 2). It may be said briefly that test 2 is "Logical Selection" (the blank presents such lists as "CHAIR—legs, rockers cushion, seat," and the children are told to underline the two things a chair is never without); test 3 is made up of problems in practical arithmetic (mostly making change); test 8 is "Disarranged Sentences;" test 9 is "Practical Information." The four tests of the arithmetic group (4, 6, 9, and 10) are opposites, "Word Completion," "Practical Information," and analogies. There is little or no "superficial relevancy," in the tests of these two groups, to English and arithmetic; but the writer does not believe that such relevancy is particularly important or to be expected. The three tests which appear of relatively little prognostic value with reference to either subject are the two memory tests and "Moral Discrimination."

for arithmetic. And a sub-scale composed of tests 2 and 8 gave 38% for English—and 7% for arithmetic. That is, there is a distinct reversal.

Such a result, if it may be considered reasonably reliable,<sup>8</sup> is surely important enough. It would indicate that prognosis problems were specific, not general. In so far as this may be true different tests and methods must be necessary for different problems. Tests of "general intelligence" can not, in such an event, be used if the most exact determination of ability (that is, the most exact prognosis) is desired; ability in English and ability in arithmetic, clerical ability and mechanical ability—each problem must be dealt with independently, and specifically studied. The point of view suggested is fundamentally different from that habitual at present; instead of dealing with specific problems on the basis of certain generalizations—dealing with them "in general"—each problem must be considered by itself, without any pre-suppositions. In attempting to prognosticate ability in mathematics, in stenography, in salesmanship, one may go through the process of considering whether "general intelligence" or a "special ability" is the important factor, may try certain tests for measuring "intelligence," and in proportion as they deal efficiently with the problem declare that "intelligence" is a factor. Why not ask instead simply the one practical and direct question "What tests will deal most efficiently with this particular problem?" And consideration as to whether "general intelligence" or this or that specific ability is involved appears largely unimportant, and in the realm of the theoretical and academic.

It is such a point of view that these results suggest. And it is a point of view that is, if one is empirically minded, very stimulating.

3. *Discussion.* Certainly the results are not such as to warrant any conclusion in the matter! But it should be kept in mind, in considering the great mass of experimentation, observation, and theory which forms the basis for current ideas regarding the dominant importance of "general intelligence," that there are certain constant errors which operate greatly to increase the apparent importance of the "common factor." In spite of the best intentions, judges tend to overestimate the similarity of an individual's abilities along different lines.<sup>9</sup> In spite of the best efforts of the test builders, it is still hardly possible to "test" much besides various types of controlled association of the verbal type, with any satisfactoriness; for any adequate sampling of the total richness of human activities the tests are still pitifully

<sup>8</sup> It is *not* reasonably reliable—of course! Both method and data are painfully inadequate to the importance of the exceedingly fundamental and difficult problem dealt with,—a problem which can hardly, in any case, be "solved," but resolution of which can come only as the result of a multitude of investigations, from a variety of standpoints, and after a period of years. The present study is simply a bit of exploratory research, and trial of methods. But as such it would seem of some interest, and suggestive value. The writer has a notion that in dealing with such large and extensive problems, a large number of such smaller studies, attacking from many angles—a campaign of attrition one might call it—are likely to be much more fruitful than a few elaborate and laborious investigations.

<sup>9</sup> See Thorndike, E. L. A Constant Error in Psychological Ratings. *Journal of Applied Psychology*, Vol. 4, 1920, pp. 25-29; also Kelley, T. L. Educational Guidance. Teachers College, 1916.

inadequate. High interrelation between tests can hardly, then, be used as evidence for close correlation between all types of ability. In spite of the great broadening of the public school curriculum in recent years, progress in school is still conditioned almost wholly by ability in the academic, bookish subjects; "general" ability, as it appears in school may be simply the functioning, with different materials, of a relatively special ability. It should also be pointed out that, while it may be convenient for practical purposes to sum the results of a series of tests into a single rating, to average school marks, or employers estimates, or otherwise develop synthetic statements as to an individual's ability, such investigations are likely to yield just what is put in by the procedure—certain vague general facts, which may very well obscure the concrete, specific details we are most in need of. The general factor may be the all-important factor. But we are surely as yet in no position to be dogmatic about it. Practical considerations may make it advisable to proceed upon that hypothesis. Meanwhile, critical investigation of the concept of "general intelligence" is surely much needed. It is as a bit of suggestive research along such lines that the present paper is intended.

A word remains to be said regarding the method which was employed, for relating tests and school record. It has obvious faults—faults so obvious that they hardly need mention.<sup>10</sup> It is simply an effort, in keeping with the fundamental idea of the study, to deal with things in the very most concrete and specific manner possible. The particular method is of little importance. But methods of this sort, which permit study of relationships in a very simple, direct, and specific way, would seem capable of rendering distinct service in dealing with many problems of applied psychology.

#### Summary

The paper deals with the scores made by a large Junior High School class, on certain tests of intelligence, and with the marks obtained by these children over a year later. The question is as to the relative merits of these tests, in foretelling what marks these children would receive. It was found that:

(1) The tests differed distinctly in their prognostic efficiency. By careful selection and combination four tests were found which were more efficient than the entire ten tests of the examination used, for the purposes studied.

(2) Certain tests appeared very efficient for certain purposes, but poor for other purposes. The four-test differential units showed this same specific efficiency; different tests appeared to be required for foretelling failure in arithmetic and English. Certain groups of tests exhibiting this specific prognosis to a more marked degree were also found.

It is urged that investigation regarding the specificity of prognosis problems is much needed.

<sup>10</sup> It is, particularly, a method which calls for a much larger number of cases, if the findings are to be reliable. It is the writer's judgment, from trial of the method in a variety of ways and on a variety of data, that it is more reliable than might be supposed. However, as an example of an analogous method yielding a more stable result see Pressey, S. L., "The Practical 'Efficiency' of a Group Scale of Intelligence," *Journal of Applied Psychology*, Vol. III, 1919, pp. 68-80, or "Suggestions regarding Professor Thurstone's 'Method of Critical Scores,'" *Journal of Educational Psychology*, December, 1919.

# MINOR STUDIES FROM THE PSYCHOLOGICAL LABORATORY OF INDIANA UNIVERSITY

## IX. FURTHER DATA WITH REGARD TO SEX DIFFERENCES

By S. L. PRESSEY and L. W. PRESSEY

*I. Problem and Materials.* In a previous paper<sup>1</sup> by one of the writers, data were presented with regard to sex differences in central tendency and variability, as they appeared on a scale of general intelligence. It was concluded that (1) the girls showed a very slight, though constant, superiority in central tendency when score on the entire examination was considered; on the individual tests they showed superiority in controlled association and memory—on 8 out of the 10 tests. The boys showed a marked superiority in arithmetic and were slightly better than the girls in general information. (2) The boys showed a slightly greater range—which was also constant for all reliable ages. Analysis by test showed a similarly greater range for the boys on seven of the ten tests; three tests showed little difference, but in no case was there greater variability among the girls. The present paper reports briefly certain further data with regard to sex differences on various types of examination.

Three groups of data are dealt with: (a) Results from 1,408 children 6, 7 and 8 years of age, tested with the Primer Scale.<sup>2</sup> (b) Results from 787 children 12 and 13 years of age, tested with the Cross-Out Tests.<sup>3</sup> (c) Results from 101 college students, tested with a series of tests intended for the "investigation of the emotions."<sup>4</sup> The reports on these three sets of data will be taken up in order.

*II. Results.* (a) The Primer Scale has been described in full elsewhere; in this connection it need only be said that it is a brief group scale for measuring general intelligence, consisting of four tests not involving ability to read or write, for use with children in the first

<sup>1</sup> Pressey, L. W. "Sex Differences Shown by 2,544 School Children on a Group Scale of Intelligence, with Special Reference to Variability." *Journal of Applied Psychology*, Vol. 2, 1918, pp. 323-340.

<sup>2</sup> Pressey, L. W. "A Brief Group Scale of Intelligence for Use in the first Three Grades." *Journal of Educational Psychology*, Vol. 10, 1919, pp. 297-308.

<sup>3</sup> Pressey, S. L. "A Brief Group Scale of Intelligence for Use in School Surveys." *Journal of Educational Psychology*, Vol. 11, 1920, pp. 89-100.

<sup>4</sup> Pressey, S. L., and Chambers O. R. "First Revision of a Group Scale Designed for Investigating the Emotions." *Journal of Applied Psychology*, Vol. 4, 1920, pp. 97-104.



three grades.<sup>5</sup> In recent surveys all the children in the first three grades of five cities have been examined—a total of 1,408 cases of which 698 were boys and 710 were girls. The results for 7 and 8 have been tabulated by sex. The six year group had to be dropped from the study because three of the five cities had yearly promotions only; the result was that in these three cities only about 60% of the six year olds were in school at the time of the surveys, and of these a majority were boys—the tendency seeming to be to keep the girls out of school and at home somewhat longer. A somewhat analogous difficulty was encountered in obtaining an unselected eight year old group; there were some children of this age who were in the fourth grade, and so not tested with this scale. However, there were only a few of these cases (8 in all, 7 boys and 1 girl) in the five cities taken together. And since there were so few accelerates, it seemed safe to assume that only the very brightest children received double promotions; these 8 children have, therefore, been counted in as above the 90 percentile for their age. With this interpolation, the data for ages seven and eight would seem fairly satisfactory.

The results, as regards central tendency, may be summarized very shortly. The girls show a slight superiority; of the children who compose the upper half of the entire distribution for these two ages, 53% are girls and 47% boys. Findings as regards variability seem to indicate a greater range for the boys. Two methods were used. The first method required simply the examination of the highest and lowest 10, 3 and 1 per cents of the entire distributions to determine the composition of the extremes as regards sex. The results appear below; the two ages have been combined and the numbers thrown into per cent. It will be seen that the boys predominate in the extremes; and the further out into the extremes the investigation is carried the more the boys are in the majority.

TABLE I

	B.	G.		B.	G.
Highest 10 per cent....	51	49	Lowest 10 per cent..	55	45
Highest 3 per cent....	81	19	Lowest 3 per cent..	69	31
Highest 1 per cent....	100	0	Lowest 1 per cent..	100	0

<sup>5</sup> See reference note 2 above, or Pressey, S. L. and L. W. "Cross-out Tests, with Suggestions as to a Group Scale of the Emotions," *Journal of Applied Psychology*, Vol. 3, 1919, pp. 138-150. The first test consists of groups of dots, each group of dots making a pattern; there is, however, one dot in each group which is outside of or spoils the pattern—this dot the children are to cross out. The second test is made up of squares, each square containing two objects which are similar, in some important way, and one object which is different from the other two; this different thing the children are to cross out. The third test shows at the top of the page four forms—a triangle, a square, a cross, and a circle. Below are squares or "boxes" containing "blocks" which are to be fitted into the four places at the top of the page; but in each "box" there is one piece which will not fit in—and this extra piece is to be crossed out. The last test shows pictures in each of which there is something wrong—this wrong part is to be crossed out.

The second method for investigating variability was even more simple. It consisted of simply examining the distributions of scores for each one of the five cities separately for ages seven and eight, in order to determine whether, in each distribution, a boy or a girl scored highest and lowest. There were, thus, ten cases of highest score (one for each of the two ages in each of the five cities) and a similar number of lowest scores. Of the 10 highest cases, every one was a boy; there was not a single girl at either age in any one of the five cities who made the highest score for her age. Of the 10 lowest cases, 6 were boys and 4 were girls.

(b) Results from three cities where complete surveys were made were included in the consideration of sex differences found with the Cross-Out Tests. After very careful consideration it was decided to limit the study to ages twelve and thirteen; because of the influence of school selection it was impossible to obtain reasonably unselected groups for any other age. As it was, a few of the twelve-year-olds and one or two of the thirteen-year-olds were still in the first three grades, where the test could not be given on account of literacy (over 10% of the eleven-year-olds were in the first three grades, while many of the fourteen-year children had left school). Since the tests included the High School, the upper ends of the twelve and thirteen year distributions were tested. So the only children of these ages who were missed were 8 retards below the fourth grade (six boys and two girls); these have been counted in as being in the lowest three per cent for their age. Of the total 787 cases, 387 were boys and 400 were girls.

The results as regards differences in central tendency agree with those found with the Primer Scale. Taking all three cities together, there were 46% boys and 54% girls scoring in the upper half of the distributions. In City A the upper half of the distribution was composed of 45% boys; in City B, 47% boys; and in City C, 47% boys.

The findings as regards variability are, however, very indeterminate. Again, the highest and lowest 10, 3 and 1 percents of the entire distributions were examined in order to determine the composition of these extremes as regards sex. The results are given below; again the two ages have been combined and the figures thrown into percents.

TABLE II

	B.	G.
Highest 10 per cent.....	48	52
Highest 3 per cent.....	43	57
Highest 1 per cent.....	50	50
Lowest 10 per cent.....	60	40
Lowest 3 per cent.....	67	33
Lowest 1 per cent (not all of the eight children below the 4th grade would score in the lowest 1%; therefore it could not be reckoned.)		

The width of the distributions for the two sexes is about the same; the table indicates little except the effect that would be expected in the extremes from the fact that the girls' central tendency is higher.<sup>a</sup>

<sup>a</sup> The scores from two of these three cities were analysed by test. The results are by no means consistent from one city to the other. Two facts, however, stand out rather sharply. On the first test, for both cities, the girls show a wider range; of the children scoring in the highest 10 per cent, 56% in City A and 63% in City B are

(c) A group of five tests "designed for the investigation of the emotions" has already been described in a previous number of this JOURNAL. (See Note 4.) Results from 101 college students, 52 girls and 49 men (students in a course in Abnormal Psychology and a course in Domestic Science, for the most part Juniors and Seniors), were analyzed by sex<sup>7</sup> and may be summarized briefly.

Test I consists of five lists, each of five words; these words are, for the most part, of emotional content,—such words as tremble, street-walker, rival, ugly, insult. The subjects were told to read through the lists and cross out *everything* that was unpleasant to them. It was found that 55% of the girls crossed out more words than the median boy. The subjects were then told to go through the lists again and draw a line around the *most* unpleasant word in each list. In scoring these responses, the word chosen as most unpleasant by the greatest number was considered the modal word for that list, and the score of each person consisted of the number of lines in which he chose some other than the modal word as most unpleasant—that is, the number of times he deviated from the modal choice of the group. (In some lines the men considered one word as most unpleasant while the girls chose another; on these cases both words were considered modal). It was found that 90% of the girls deviated more than did the median man.<sup>8</sup>

In another test, in which the subjects were required to read through 25 lists of moral terms (such as insulting, quarreling, mislead, hurting, careflessness), and cross out the worst thing in each list, 22% of the girls deviated from the modal choice more than the median man. Similarly, in going through the same lists again and selecting the most common sin, 26% of the girls deviated more than the median man.

In a third test, in which the subjects were required to read through 25 lists such as FOOT—leaf, stocking, tennis, spider, kick, and cross out the word in small letters most closely connected in their mind with

---

girls; of those scoring in the lowest 10 per cent, 69% in City A and 63% in City B are girls. The other fact is a similarly greater range for the boys on the third test. Of those scoring in the highest 10 per cent, 59% in City A and 67% in City B were boys; of those in the lowest 10 per cent, 56% in City A and 58% in City B were boys. It should be said at once, however, with regard to this last result, that on a test somewhat similar to this (another arithmetic test) used in the previous study referred to in Note I, the range for the boys and the girls was identical, although the central tendencies were, in both cases, higher for the boys. Test I, it should be mentioned, is a form of Disarranged Sentences; Test II is a test of practical information; Test III is a form of number progression; and Test IV essentially a vocabulary test. On tests two and four there appeared to be little difference in variability, or in central tendency.

<sup>7</sup>The tabulations were made by Mr. O. R. Chambers, to whom the writers wish to express their obligations.

<sup>8</sup>It should also be mentioned that the words of the test were divided into four classes according as they were unpleasant because they appealed to fear, disgust, sex or self-feeling. Half again more sex words were selected by the girls than by the men as most unpleasant, while half again more boys than girls selected the self-feeling words as most unpleasant. The men exceeded the women in the selection of fear words.

the word in capitals at the beginning of the list, only 27% of the girls showed idiosyncrasy greater than the median man.

These results are, of course, from only a few selected, adult cases, and are results on only three tests.<sup>9</sup> They are included in this report because of certain important bearings such findings have, the writers believe, upon the interpretation of results obtained with tests of intelligence.

3. *Discussion.* The above materials surely make up an odd collection of results, from which it would hardly seem possible to draw any very definite conclusions! It seems fairly clear that on the usual scale for measuring "general intelligence" girls average slightly better than boys. The differences are, however, so slight as to be essentially negligible from any practical point of view. In fact, the slight superiority of the girls may very well be due simply to a preponderance of tests of a literary nature, in such scales, as they are usually made up. The boys show superiority on other types of tests—as has well been demonstrated by use of the army tests in the universities. In fact, the writers would be willing to contract to build a group scale of "intelligence" in which the boys would excel. Or, perhaps the fundamental fact in this superiority of the girls is dependent in part upon a greater docility or upon a relatively greater interest in indoor tasks. Or, the difference may be connected with differences in physiological growth rate—one might enumerate almost any number of possible causes, but surely with no great profit. The one fundamental and all-important fact about these differences in central tendency is, as Thorndike says,<sup>10</sup> "their small amount."

As regards variability the situation is much more complex. The findings obtained by the Primer Scale are essentially in accordance with the results obtained previously by one of the writers (see note 1). The findings with the cross-out tests fail to support these conclusions. It may be that the more narrow distribution of scores given by these tests does not give room for the differences to show themselves. There is no good reason that the writers can see, however, for supposing this; the amount of it is that the total findings are not consistent with each other. Certain considerations with regard to what positive findings there are, indicative of greater variability among the boys, remain, however, to be mentioned.

It should be pointed out, first of all, that the positive findings, even if they be accepted at their face value as indicative of fundamental greater variability in mental calibre among boys and men, do not indicate differences in variability sufficient to warrant any practical action in educational policy, in business, politics, or elsewhere. It should also be made clear that the differences between the most brilliant boy and the most brilliant girl are so slight as to make it altogether unwarrantable in particular instances to discriminate toward or against either sex. The positive findings, even with the most that can be made of them, have no more than a theoretical and general interest.

But there are a number of other entirely plausible explanations which can be given for the positive findings as regards variability—explanations which do not presuppose any difference in the distribution of intelligence. Three such explanations seem to the writers particularly worth mention.

<sup>9</sup> Two other tests were given (see the original paper, note 4) but the results yielded were of doubtful reliability.

<sup>10</sup> Thorndike, *Educational Psychology*, Vol. 3, 1914.

(a) It has already been pointed out that the higher average of the girls, in total score on the scales used, may be due simply to the nature of the tests making up these scales—to a more or less chance of selection of tests on which the girls did better; and the writers have ventured the assertion that it would be entirely possible to formulate a scale showing a similar slight superiority for the boys. There is a little evidence to suggest that, in analogous fashion, comparative variability may be conditioned by the nature of the tests (as with tests I and III of the Cross-out Scale).

(b) It must also be emphasized that selective influences, of an extremely subtle and pervasive kind, operate to make it exceedingly difficult to obtain sex groups so thoroughly unselected as to make suitable material for the study of nice differences in variability. The difficulties in this connection encountered in the present study may serve as illustration.<sup>11</sup> The possible importance of differences in age of entering remains, however, to be pointed out. In certain of the systems examined the difference was marked; thus in City G, there were 63 boys six years of age and only 46 girls. If we may suppose that the boys, as a result of circumstances, tend to enter school earlier, but for the most part drop back; and if we may suppose the tests to be a slight degree affected by school training, we evidently have a very pretty explanation for greater variability on the part of the boys. They have been in school longer, the brightest boys are further along than the brightest girls; but, for the most part, they are less developed, or less interested in school work, and the dullest ones soon drop behind the dullest girls. This explanation is, the writers feel, of some little interest.

(c) It should, finally, be emphasized that there are very marked differences between the sexes in emotional attitudes and interests—differences which are the most marked findings in previous studies, and which have received interesting confirmation in the emotional test data just presented. Such differences may be the result of differences in instinctive equipment. They are surely also, in no small part, due to social influences. It would seem not at all unlikely that such factors might have their influence upon results yielded by tests of "intelligence." Whether they could operate to affect variability may be a question; but several ways are conceivable in which they might bring about such a result. In considering any differences between the sexes, emotional elements would seem deserving of first consideration.

Well—surely no very definite conclusions are possible! In fact, the writers have come to feel that the test instruments are hardly at present developed to the nicety required for satisfactory study of sex differences, so far at least as variability is concerned. And they have largely lost interest in the problem. They are, in fact, rather inclined to feel that it might be wiser to speak of tests of intelligence as tests of general prognosis, as is suggested in the accompanying study, without attempting to decide as to just what factors may contribute to the prognostic indication. It then at once appears that such tests are hardly suited for investigation of this problem of sex differences in mental endowment.

<sup>11</sup> See also Pressey, S. J., "The Problem of the Unselected Group in the Standardization of Tests," *Journal of Applied Psychology*.

## SUMMARY

The article summarizes certain data with regard to sex differences in mental and emotional traits, as determined by various tests. It was found that:

1. Results with the Primer Scale show a slight superiority in central tendency for the girls and slightly greater range for the boys.

2. Results with the Cross-out Tests show a slightly superior central tendency for girls; the findings with regard to variability are indeterminate.

3. Results with a series of tests "designed for the investigation of the emotions" show definite sex differences among college students in likes and dislikes, moral judgments and emotional associates.

4. It is concluded (a) that such sex differences in intelligence are of no practical importance, (b) that it is hardly possible at present to investigate these differences with any degree of exactness because of the difficulty in calculating the influence of (1) the subject-matter of the test, (2) selective factors, and (3) emotional factors, probably of environmental origin.

## A NEW JOURNAL OF PRACTICAL PSYCHOLOGY

### And a Translation of Its Introductory Article

By HENRY C. LINK

In October, 1919, Germany saw the appearance of a new monthly journal devoted to the field of practical psychology. This journal, entitled "*Praktische Psychologie*," is edited by Dr. W. Moede and Dr. C. Piorkowski (Berlin W. 30, Luitpoldstr. 14). The former is a teacher at the Charlottenburg University and is carrying on extensive experiments with tool apprentices. The contents of the first two numbers of this journal have already been briefly reviewed by the writer in the *Psychological Bulletin* for October.

In introducing this journal, the editors say in part:

"*Praktische Psychologie*" is the first monthly paper which is conducted in the interests of the entire field of applied or practical psychology. It will concern itself chiefly with industrial psychology and experimental pedagogy, but also with medical, legal and social psychology as well as with experimental esthetics. The chief emphasis of the publication, however, will be upon Industrial Psychology.

The field of industrial psychology includes in the first place:

"Vocational aptitudes and vocational guidance. The tremendous importance to the new state of this particular subject is indicated by the appropriation for vocational guidance made jointly on March 19, 1919, by the ministry of industry and commerce, the ministry of the interior, of education and of agriculture. Vocational guidance is thereby made a duty of the local community, and the careful consideration of bodily and mental aptitudes is strictly enjoined. We shall discuss impartially all statistical and experimental studies which contribute to the knowledge of the necessary facts regarding a particular vocation."

"A second important aspect of Industrial Psychology is the reorganization along more rational lines of the apprenticeship system and of working methods. These subjects will be thoroughly discussed as a part of Industrial Psychology. Above all we shall here describe the steps by which operations at a machine or at a desk are to be improved in the direction of increased comfort or decreased fatigue.

"In addition to actual operations we shall also thoroughly consider questions of external and internal organization in the light of psychological technique. The most effective methods of instruction will receive our particular attention."

The first article in this new journal is an eloquent exposition of the significance of the new science for the future of Germany. It was thought that American psychologists would be interested in this point of view and in the scope of science as it is here described. For this reason, permission was obtained from Dr. G. Schlessinger, the writer, and from the editors, to translate and publish the article in this country.

## VOCATIONAL GUIDANCE AND PSYCHO-TECHNICS

Professor ING G. SCHLESSINGER at Charlottenburg

The most powerful stimulus of the German revolution is the recognition of the fact that the destinies of the world are determined not by matter and its organization, not by force and its application (in these respects the utmost was accomplished during the war) but by the human being in the capacity of manager or executive, leader or subordinate. The exaggerated mechanization of human beings, compulsory labor, the censorship, excessive discipline, led inevitably to rebellion, to the violent breaking of the chains which bound a newly unified people. Man is not a machine. There is no money equivalent for personal sacrifice. Therefore, our workers today are struggling not for higher wages, merely, but for moral values. Above all, they wish to elevate themselves from the degraded level of the fourth class to the level of equal rights and privileges. "We workers wish to share in knowledge (mitwissen) and to cooperate (mitwirken) where practicable, in management. Wages, however, shall be based on accomplishment."

"Granted that the humanity and the personal dignity of man now hold the center of attention, and that man-hours and man-units are no longer to be treated as commodities, the supreme problem before us is to place the best man always in the best position—in the trade school, where the apprentice develops, in industry where laborer and executive are employed, in the development department where machines are invented and prepared for the manufacturer, in the executive department to which the threads of industry lead and in which they must be kept free from entanglements and undue tension. To find the right man and place him in the right position is today more than ever, the supreme problem in our broken down Fatherland, for although raw materials are scarce, leaders are even more scarce.

While we are waiting for a Messiah, some conspicuous leader who will help us out of our deep distress, we cannot remain idle. We must give rein to our old German virtue, "Strenuous and systematic work." We must exert ourselves to find or create the means which will infuse new strength into Germany's mainstay, namely, its powerful Industry. We must cooperate in the laying of that human foundation without which our recovery is impossible.

Enthusiasm and inclination are the conditions of great deeds! But who can work energetically and cheerfully if he has chosen the wrong vocation? Who can help a youth, when he leaves school and lacks a well-based preference for a particular calling—an occurrence which is all too frequent—to find the right path? The father is onesidedly absorbed in his own occupation, with which he himself is frequently dissatisfied. The teacher usually knows only superficially the requirements of the various activities about which he is called upon to advise his pupil. He can advise him only morally and ethically. The physician can, on the basis of a few physical examinations, only inform his patient whether his health and strength will suffice for a particular vocation.

Therefore, we are confronted at this critical point by a gap, which, if not bridged, will have the most serious results on the working effectiveness of the nation.

The choice of vocation and thorough vocational training are bound up with effective and practical vocational guidance. The latter requires experts whose knowledge must be accompanied not only by



honest intentions and a warm heart (who can call himself a vocational counsellor without these?) but by a mastery of the entire technique which enables us to illuminate the mental makeup of the person to be advised. Though many of the mental activities of man are only mechanical, how fallacious it is to rate copying, counting, and figuring higher, mentally, than driving an automobile, cutting a micrometer screw, or filing a fine piece of work. These are only different mental activities." At one kind of work hands remain clean and collars white, but the degree of concentration, the feeling of responsibility are in no way more intense and are only necessary to a greater extent.

The science of management, therefore, rightly gives first emphasis to the proper selection of men. It always begins—if it is to be successful in the central office of an industry, among the so-called brain workers. It appoints these, and not only lays out the plan of operation, but above all designates those who are to execute it. It ends only in the shop with the actual process of manufacturing. Its influence is strongest and most penetrating from the top down rather than in the reverse direction. With a practical man one can talk business. Every word starts a chain of thought, releases a series of experiences, with which it is easy to tie up new ideas. To convince a routine encrusted, academic bookworm of the value of new methods is, as experience tells us, much more difficult because the powerful mental influence which actual work in a shop exerts is missing and must be replaced by abstract education through the spoken word and the use of general concepts.

Therefore, the content of vocational guidance—Psychotechnics—is based for the present, above all things, on the discovery of the mental abilities of the apprentice. A technical procedure is required in order to analyze the inherited tendencies of an individual though not to explore his soul. The word "Psycho" is, to be sure, misleading; our concern is not at all with the "soul" or with morals or ethics, or with a man's will-power; the latter can only be discovered through personal association, and for this the teacher in school is the proper person. From the standpoint of the practical psychologist we are concerned only with the *ostensibly* spiritual. Those objective manifestations which the individual himself can verify and which will offend no man. For instance, whether an arm is strong or weak, an eye far or near-sighted or color-blind; whether the mathematical sense is crippled, the memory weak, or the susceptibility to fatigue marked—these are determinable facts which will never embarrass an individual. We are concerned not with demotion but with proper classification. For the presence of a number of these attributes, partly physical, partly mental, is indispensable to the successful exercise of a particular vocation. The blacksmith needs a strong hand, the locomotive engineer must be able to distinguish colors, the piano tuner must have a musical ear, a librarian a good memory, as must also a writer. The pilot of a boat, an airship, a motorcar, must not become easily exhausted.

The dilettante scientist is therefore unable to study psycho-technics as an abstract science. He must start with a particular vocation, the elementary requirements of which he must carefully gather, sift, and analyze before he can undertake to apply his tests with any chance of success. This preparation the psycho-technical procedure prescribes as indispensable. Then follows the search for apparatus with which to measure these innate qualities, and finally the construction of the

apparatus. The apparatus can be borrowed from the immediate requirements of the vocation itself—for example, a micrometer as a test of the sense of touch; it may, however, be quite independent of such a connection, as, for example, the tremometer for measuring tremor in core-makers, the tachistoscope for street-car motormen, etc. Slowly, but surely, the close cooperation between engineer and business psychologist which has now been achieved is developing a procedure which will apply equally to all vocations, and which will vary only as to the apparatus required by the different occupations. The machine designer and the mason, the automobile driver and telephone operator have more mental activities in common than the superficial observer generally thinks. And yet the entire program of the industrial psychologist is far removed from any attempt to expose the soul, a belief with which ignorant opponents try to frighten the workers.

One who has seen the many hundreds of fourteen to sixteen-year-old apprentices at the Charlottenburg Technical High School, working cheerfully and even enthusiastically in the vocational experiments being conducted there for machine tool and shop management work, a group which Dr. Moede is handling, one who has lived through the tragic circumstances which called attention to the necessity for such experiments, will be convinced that we are on the right track, that we are indeed bridging the gap which has hitherto separated the school life from the vocational calling, and that we will open our accounts on a fresh page when parental home and the school, medicine and industrial psychology, cooperate toward a common end.

Neither science nor experience are the determining factors nor are we primarily concerned with calculations, inventions, or experiments. Our chief concern is the full, unhindered development of inherited qualities, in short, with the really best development of every single being. Only in this fashion can we convert workers, from the manual laborer to the general manager, into skilled workers. And only thus can we attain the high level of humanity from which the longed for leader will finally arise.

## BOOK REVIEWS

J. CROSBY CHAPMAN. *Trade Tests*. Henry Holt and Company, New York, 1921, p. 435.

As is indicated by its subtitle, this book by Professor Chapman constitutes a manual on "The Scientific Measurement of Trade Proficiency," and answers the general question, "What is a trade test?"

The chapter headings which follow are indicative of the scope and character of the work: The Construction of the Oral Trade Test; Administration and Scope of the Oral Trade Test; The Picture Trade Test Method; The Construction of a Performance Trade Test; The Written Group Test Method; The Analogy between the Army and the Industrial Employment Problem; Problem of Occupational Analysis; What the Trade Test is Not; The Use of the Various Types of Trade Tests; The Adaptation of Trade Test Methods to the Training of Employees; the Installation of Trade Tests; and Some Further Applications of Trade Test Methods.

The reader is led to expect much from so imposing a list of contents, and his anticipations are fully realized in reading the book, for a clear-cut presentation is made of the scientific material collected. This is further supplemented by many illustrations of trade tests, rating scales, diagrams, photographs, etc.

"Trade Tests" are a by-product of the World War. They are the result of extensive scientific research work conducted by the Committee on Classification of Personnel under the direction of the War Department to solve the problem of efficient selection and placement of the army's skilled personnel. In view of the extraordinary success of the trade test instruments in the army, and having in mind the utter inadequacy of the methods employed by industry at large, the author presents the findings as a method which, in a modified form, should be extremely useful to industry in solving its perplexing problem of securing an adequate amount of the indispensable human element in the process of production.

Professor Chapman had a direct part in formulating the trade tests, and he gives a clear and comprehensive statement in this volume of the conditions which gave rise to these tests, he discusses the methods of their construction, their relative merits and demerits, and he suggests their various possible applications to the problems of the industrial world with reference to its personnel. He sees an analogy between the army and the problem of industrial employment, and believes that by adapting these trade tests to the needs of employment bureaus and vestibule schools, industry will get much nearer the solution of its problem.

Under present slack conditions with great masses of workers idly twiddling their thumbs, the idea of introducing a labor saving device appears, to say the least, somewhat ill-timed. One fears that the statement that "depleted man-power is the sad aftermath of war," under present conditions, will be far from conducive to the wide circulation which the book merits. Dr. Chapman evidently considers the present ebb in industry but a temporary condition, and anticipates little diffi-

culty from this direction. He fears that opposition may come from the man for whom the book is primarily designed,—from the employer who has had sad experiences with unqualified individuals who have offered various inadequate solutions in the past, and that he may see in this extremely scientific method but another bogus cure.

The work is written entirely from the viewpoint of the employer's needs, emphasizing the benefits which the trade tests should yield to him; and one is disappointed that the author has not pointed out the advantages that should accrue to the workers and to society in general by the application of these tests.

Veblen has pointed out the fact that the introduction of machinery has reduced everything to standardized units, i.e., not only production processes in industry, but man's working conditions, his living conditions, the goods he consumes, and his very needs. A disturbance anywhere of this standardization means hardships. Since the human element is so important, which the author recognizes, he offers these trade tests as a means of stabilizing this fluctuating human factor.

Labor should be quick to see the advantages it would gain by the extensive use of this method of gradation and selection. It should result in removing round pegs from square holes. Placing a man on the job for which he is fitted by nature, inclination and training, and where he will have opportunity to express himself to the utmost, should strengthen what such students as Parker, Veblen, Tead, and others, include in the general phrase of "instincts of workmanship." The ideal consequences should be greater production for the entrepreneur, advancement and increased pay and contentment for the workman, and a step toward industrial peace for society at large.

The author is looking forward to the day when various scientific methods will be employed by private and public employment bureaus where by means of an adequate examination the man seeking work will be given his trade rating and sent to the right job. He urges the imperative need of a more economical use of industry's great want, man power.

"Trade Tests" is the result of much painstaking effort by such distinguished men as L. E. Thorndike, W. D. Scott, the author, and others. All students of industrial problems will find in this scientific work a distinctly new method which, in connection with intelligence tests and skill prediction tests, gives promise of effectively applying psychology still further to the far-reaching problems of our national industrial life. The book cannot be too highly commended. It should receive the widest circulation.

Clark University.

G. T. SCHWENNING.

KNIGHT DUNLAP. *"Mysticism, Freudianism and Scientific Psychology."* C. V. Mosby Co., St. Louis, Mo.

This book may readily be classified as being of two parts: destructive and constructive. Professor Dunlap has little use for Mysticism and Freudianism, so the first two chapters on these subjects are destructive in nature, while the third chapter or constructive part deals with the foundations of Scientific Psychology.

Taking each term separately, the author devotes a chapter to each in a clean-cut description. Mysticism is defined as a belief in a third kind of knowledge which ranks higher in value than either of the other two. There are a number of conventional names, all of which are admitted to be metaphorical, for the act or fact of knowing in this assumed third way. Three terms especially have been employed to

designate this act: "union," which emphasizes the fundamental characteristic that the subject who knows and object which is known become the same by one being absorbed into the other; "love," that which attracts and binds together; and "ecstasy," which alludes to the same alleged fact of the knowing subject losing its identity in the known object. After discussing the origins of the European Mysticism, Quasi-Mysticism and Pseudo-Mysticism, Professor Dunlap has this to say in the closing of the chapter: "In freeing itself from the arms and methods of science, Mysticism adopts without scruple a type of reasoning against which science constantly struggles, the type known to the logicians as the fallacy of the ambiguous middle term or the 'sliding term.'"

The second chapter discusses Freudianism and the psychoanalysts, giving the origin, program and claims of psychoanalysis, and setting forth the essential postulate of the "subconscious mind." The most important ideas or furniture of the mind are "desires," the most important being sex desire which is assumed by Freudians to commence in the early weeks of infancy as autoerotism. This hypothesis of infantile sex life is founded on the specific fallacy of "secundum quid." The final consequence of this method is to define sex as the whole universe, which would necessitate the derivation of a new term for what is ordinarily considered sexual. Freudianism, like Mysticism, is also unscientific divulging into the fallacy of the ambiguous middle term and venturing into a realm resembling that of the third kind of knowledge of Mysticism.

Anything to be scientific must be, like Scientific Psychology, founded upon certain scientific procedure. Science must start from an empirical basis of the facts of experience; must form, according to the law of parsimony, a working hypothesis, subject to experimental investigation, into which the observed facts must fit; and must pay extreme care to significance of terms. Such principles readily exclude at once the dangers of Mysticism and pave the way for a genuine physiological psychology or psychobiology, which is portrayed in the third chapter.

While being opposed to the Freudian movement, the author asserts the opinion, however, that its final result may be beneficial. Just as Christian Science has tremendously accelerated the progress of scientific medicine, so psychoanalysis, by compelling psychology to put its own house in order, will eventually help in the development of Scientific Psychology. But in the fear that in the meantime many may be deluded into the "short-cut" methods of Mysticism and Freudianism rather than adhering to the slower but more solid procedure of science, this book was written.

Clark University.

CLINTON S. LEONARD.

W. WHATELY SMITH. *"The Foundation of Spiritualism."* E. P. Dutton & Company, New York, 1920, p. 123.

This book is a discussion of the fundamental bases of spiritualism. The author examines the various phenomena which serve as a foundation upon which the claims of spiritualists are built.

A large part of the book is taken up in analyzing the spiritualistic phenomena which are used as evidence for survival and gives some instances of automatic writing which seem to show that departed spirits take some interest in, and maintain some measure of contact with, the world they have left. He also takes up the arguments against the spiritistic hypothesis and says that "the true case against the

spiritistic hypothesis is simply that our knowledge of abnormal psychology is not, at present, extensive enough to admit of our assigning definite limits to what can be achieved by the incarnate mind under suitable conditions, and that, until we are in a position to do so with reasonable precision, we cannot possibly say that any particular feat is beyond its powers and therefore necessarily due to the interventions of discarnate intelligences." Part II of the book has to do with "The Process of Communication" and the author concludes that the whole subject is so obscure "that it is at present premature to express any opinion whatever as to the nature of the next life." He does, however, suggest a few conclusions as to the nature of the individual in the next life. The author is somewhat shaky in his conclusions and admits that a great deal of the evidence is unreliable and perhaps only about 5% is dependable. It is hard therefore to sift the reliable evidence from the unreliable. He states that wholesale denial of the facts of spiritualism merely displays ignorance of the subject and concludes that "the subject is one which should be dealt with exclusively by the expert and not the amateur," and that spiritualism is to be deprecated "because of the scientific immaturity of the whole subject, the great difficulties connected with it, the unfortunate tendency of Spiritualists to indulge in unwarrantable, and sometimes heterodox extrapolations from the facts, and the ever-present danger that their uninstructed adherents accept as inspired revelation matter which is really no more than the product of incarnate subliminal activities."

The author's style is clear, the method analytic. He assumes a critical attitude throughout and weighs carefully the evidence in favor of and against survival. The book is a sane analysis of spiritistic phenomena and should prove helpful as a careful and critical review to all interested in Spiritualism.

M. MEENES.

*Courtis Standard Practice Tests in Handwriting.* By STUART A. COURTIS and LENA A. SHAW. Specimen Set, 50 cents postpaid. Yonkers-on-Hudson, New York: World Book Company.

This material, like the Courtis Standard Practice Tests in Arithmetic, is thoroughly standardized. Three years' trial in schools, before it was placed on the market, left no doubt on its success in use.

Convinced that writing is a trick which the pupil must learn for himself, though his teacher may help him at times, the authors have placed before the child definite attainable goals, based on standards, and have provided exercises that enable him to reach those goals. Through these exercises, the individual needs of the children are met and the routine of classroom work is avoided. Children learn quickly how to use the material, and enjoy using it.

Standard research and supervisory tests are included in the material; by their use it is easy to discover just what kind of drill work children need, and how much.

The results from using the tests are very remarkable, many instances showing that even with less practice time than usual, children doubled their speed and quality of writing when they had used the Courtis Standard Practice Tests in Handwriting.

Bulletin No. 1 gives complete information on the material.

ORDWAY TEAD and HENRY C. METCALF. *Personnel Administration*. McGraw-Hill Book Company, New York, 1920. pp. 538.

In "Personnel Administration" the authors are offering the public a most timely book. While they may appear to be somewhat premature

in setting forth in a systematized manner the "principles" and "standard practices" of a field so new and unformed as that of the administration of human relations in industry, it is in just this respect that the authors are performing an extremely useful service. They have taken the bewildering mass of written material relative to this extensive topic and have subjected it to a critical examination, careful selection and classification before putting it in book form, and so have answered a definite need. It is in this work of expert organization and sifting of the sources rather than in the originality of the data presented in which the writers of this volume have made their greatest contribution. An additional feature which makes the book valuable is the voluminous and excellently classified bibliography given at the close of each chapter.

The volume will be useful as a handbook as well as a text for class purposes, and will have value to all "who are interested to advance right human relations in industry," be they social welfare workers, students of industrial problems, the workmen themselves, consumers or the managers. A clear, concise discussion is given of the various elements that constitute the science of personnel administration. The more important of the numerous methods that are being successfully employed by progressive industries are related, and significant generalizations are drawn from this material. The book does not propose to offer a panacea for all industrial ills. It does claim, and the claim seems to be justified, to contain a method that is working in the right direction, and bases its claim on the success which has attended the application of these principles wherever they have been properly applied.

In the estimation of the reviewer, the authors wisely urge the point that the problem of directing human beings is a permanent one, i.e., that it will remain wherever and so long as there exists the "relationship of employer and employed, of manager and managed," and this entirely independent "of the problem of ownership in industry." So long as this dual relationship in industry exists—and this seems to be *ad infinitum*—the problems involved in engaging men and in directing their activities while at work will remain no matter how often the ownership of industry changes. If the problem is inherent in industrial society, it is clear that a complete solution cannot well be devised over night, and that a wider and more systematic study of the underlying principles that will make for as little friction as possible in our industrial organization is of great importance.

Chapter II on "Human Values in Industry" reiterates, ably though only briefly, yet again the need of reckoning with the indispensable human factor without which no concern can long exist. The chapter is based on the fundamental human instincts and urges anew that they are found universally in the managed as well as in the managing, a consideration that has too long been neglected with unfortunate results as a consequence. To be sure, "it is manifestly true that in a majority of corporations production is to-day affected adversely not so much because of technical inadequacies as because of the failure of managers to recognize that workers are human beings who demand the considerate treatment which only intelligence and insight regarding human nature can suggest."

In reading this book one is impressed with the bigness of the task of personality direction and with the tremendously large, delicate and intricate machinery it necessitates creating to successfully man-

age this human section of modern industry. The nature of the problems involved, their implication to the welfare of society at large, and their bearing upon the productiveness of industry, these elements elevate the profession of personnel administration to a high plane and first rank in industry. This point is stressed by the authors calling attention to the fact that upon sympathetic interest, mature judgment and adequate preparation in the social sciences depends the successful direction of so delicate and responsible a task as that involved in this calling.

All those interested in producing and maintaining cordial human relations in the industrial world will welcome this excellent text on personnel administration.

Clark University

G. T. SCHWENNING.

DAVID E. BERG. *Personality culture by College Faculties*. Institute for Public Service, N. Y. City, cloth, 128 pp. \$1.50.

This is a study of 72 university instructors in 100 classes in 25 subjects. A summer was devoted to visiting the classes. A description of each teacher as seen from the viewpoint of the students is given. Early observations of eight instructors revealed the examples of personalities ranging from the "zenith to the nadir" in teaching ability,—the poor teaching power being due to correctable faults of personality." With each group of personality portraits from the "lower levels" in personality to the "personalities plus," is given the outstanding personality elements that determine the classification. Each picture is so fully described as to show clearly those qualities of personality that should be cultivated and those avoided.

Of a total of twenty-two there were eight of high class personality, six medium and eight of defective personality.

Some of the desirable volitional elements listed are "exactness, moral courage, aggressiveness, decisiveness, dynamic" and among the opposite are "laxity, subservience, timidity, irresolution, discouraging." The 72 instructors as the result of analysis and classification fall into 10 grades of teaching ability. It is the hope of the author that this attempt to distinguish these various grades of teaching ability may stimulate teachers toward self-analysis of their own ability and encourage colleges to consider more comprehensively the subject of teaching personality.

Every college is urged to organize a committee to study teaching personality and efficiency and give the committee power to study, analyze and formulate recommendations for improvement, that this committee formulate minimum standards of teaching personality, and study their own pedagogical problems.

This book has a fact basis for its content, and its concreteness will enable it to be of valuable service to present teachers and teachers-to-be.



## NOTE

### FUNDS FOR SCIENTIFIC RESEARCH

The Research Information Service of the National Research Council has recently compiled information about funds for scientific research. From this compilation it appears that there are hundreds of special funds, trusts, or foundations for the encouragement or support of research, in the mathematical, physical and biological sciences, and their applications in engineering, medicine, agriculture and other useful arts. The income from these funds, which amounts annually to at least fifty million dollars, is used principally for prizes, medals, research scholarships and fellowships, grants and sustaining appropriations or endowments.

So numerous have been the requests to the Research Council for information about sources of research funds, availability of support for specific projects and mode of administration of particular trusts or foundations, that the Research Information Service has created a special file which it is proposed to keep up to date in order to answer the questions of those interested in such funds. Furthermore, in order to give wider publicity to the immediately available information about research funds, the Council has issued a bulletin under the title "*Funds available in 1920 in the United States of America for the encouragement of scientific research.*"

Inquiries concerning the bulletin or for information about research funds should be addressed, National Research Council, Information Service, 1701 Massachusetts Avenue, Washington, D. C.

---

The following books and pamphlets have been received:\*

Army Lessons in English (7 copies). School Office, Recruit Educational Center, Camp Upton, N. Y.

Bureau of Business Research, New York University. Attention Value of Advertisements. The Graduate School of Business Administration of New York University, November, 1920. pp. 32.

Commissioner of Education. A Survey of Education in Hawaii. Washington Government Printing Office, 1920. pp. 408.

Education for Citizenship. Washington Government Printing Office, 1921. pp. 30.

Education for Highway Engineering and Highway Transport. Dept. of the Interior. Washington Government Printing Office, 1921. pp. 134.

Mabel R. Fernald, Mary H. S. Hayes and Almena Dawley. A Study of Women Delinquents in N. Y. State. The Century Co., 353 Fourth Ave., New York City, 1920. pp. 542.

E. H. Fish. How to Manage Men. The Engineering Magazine Co., New York, 1920. pp. 337.

---

\* Mention here does not preclude further comment

- Edward A. Fitzpatrick. *Genesis and Purpose of Vocational School Survey, Beginnings of Continuation Schools in Wisconsin, History of Continuation Schools in Wisconsin*. State Board of Education, Madison, Wisconsin, 1921. pp. 23.
- Capt. J. M. Hanson (Editor). *The Inter-Allied Games*. Games Committee, 1919. pp. 511.
- Prof. G. B. Hotchkiss, R. B. Franken. *Newspaper Reading Habits of College Students*. Association of National Advertisers, Inc., New York City, pp. 15.
- Hoyt and Peet. *Everyday Arithmetic, Primary*. Houghton Mifflin Co., Boston, New York, Chicago, 1920. pp. 276.
- Hoyt and Peet. *Everyday Arithmetic, Intermediate*. Houghton Mifflin Co., Boston, New York, Chicago, 1920. pp. 276.
- Hoyt and Peet. *Everyday Arithmetic, Advanced*. Houghton Mifflin Co., Boston, New York, Chicago, 1920. pp. 324.
- Principles, Plans and Purposes of the Educational Program of the Recruit Educational Center, Camp Upton, N. Y., 1920*. pp. 19.
- Luella Cole Pressey. *The Relation of Intelligence to Achievement in the Second Grade*. From *Bulletin of Extension Division, Indiana University*, Vol. VI, No. 1. pp. 80.
- National Intelligence Tests*. World Book Co., New York, 1920. pp. 32.
- Carl E. Seashore. *A Survey of Musical Talent in the Public Schools*. University of Iowa, Iowa City, 1920, pp. 36.

# THE JOURNAL OF APPLIED PSYCHOLOGY

VOL. V

JUNE, 1921

No. 2

## THE MEASUREMENT OF AGGRESSIVENESS

By H. T. MOORE and A. R. GILLILAND, Dartmouth College

The term aggressiveness, as here used, is synonymous with personal force, initiative, assurance. It is thus understood as standing for that trait which in combination with intelligence and reliability goes far towards completing the essential personal requisites for success. Translated into the terminology of McDougall it is equivalent to an abundance of self-assertion and pugnacity, and a minimum of fear. Described behavioristically it is invariably contrasted with the attitude of inferiority or submission.

While we may never find exactly the same pattern of aggressive behavior in any two individuals, there is nevertheless a widespread and reasonably established belief that one generally knows what to expect of the aggressive man in most of the ordinary circumstances of life. He is more likely to be vigorous, positive, and masterful than the man lacking in this trait; and he is less likely to shrink from notice, to avoid argument, to display a lack of "nerve." Generally speaking his chances of success in any type of undertaking are higher than those of the average man. Undue aggressiveness may handicap personal progress in a certain number of cases, but it is probably not so much the excess of personal force as the lack of tact and sympathy that is the drawback in such cases. So we may say with only slight qualification that, other things being equal, the measure of a man's aggressiveness is the measure of his chance of success.

The laboratory tests of aggressiveness here reported are to some extent the outgrowth of the instinct-emotion tests reported previously in the *American Journal of Psychology*.<sup>1</sup>

<sup>1</sup>"A Method of Testing the Strength of Instincts." *A. J. P.* April 1916, 227-33; and "Laboratory Tests of Fear, Anger and Sex Interest." *A. J. P.* July 1917, 390-5.

But the method of attacking the problem is fundamentally different. The first step in the present experiment was to find two groups of subjects who contrasted to a marked degree in the trait in question. This condition was met by taking from a class of 89 students the highest and lowest fifteen per cent as rated on aggressiveness. This gave thirteen men in each group. Reliable ratings of these men were obtained only after much labor, but it is believed that the method used is quite satisfactory for the present purpose, namely that of guaranteeing a marked difference in aggressiveness between the highest and lowest groups.

The rating of each man was determined by averaging two types of estimates, those made by the faculty and those made by fellow students. In order to secure standard faculty ratings the department of psychology proposed to the faculty a general system of personal ratings of the students. The plan which was adopted is similar in many respects to the officers' ratings used in the army. At the end of the college year each instructor files in the office of the Registrar a rating sheet containing his personal estimates of the men under his instruction. The estimates are made on a scale of five letter grades, as in the case of scholarship marks, and the traits judged are intelligence, aggressiveness, reliability and personality, all of them carefully defined on the rating sheets provided for each instructor. To offset the possibility of perfunctory and therefore careless ratings the point is stressed that no instructor shall under any circumstances submit any ratings which fail to represent a reasonable amount of confidence on his part. It is urged that where it is impossible for a real judgment to be formed, a blank is to be preferred to a forced opinion. Conformably to the above plan the Dartmouth faculty submitted in June 1920 personal data concerning 1,480 students. Practically all the men in the class of 89 students from whom the subjects for the experiment were to be taken were rated by from three to five of their instructors.

Student ratings were similarly obtained from the members of the class. Each student was asked to rate those of his fellow students whom he knew particularly well. From the averages of the faculty and student ratings the list of the thirteen most aggressive and the thirteen least aggressive men was now determined. After making all possible allowance for individual inaccuracies of rating, it seemed absolutely certain that the thirteen highest men, taken as a group, were conspicuously more aggressive for any purpose whatsoever than the lowest thirteen men. As corroborative of this it is

worth mentioning that the thirteen highest men included the president of the senior class, the business manager of the college daily, the manager of track athletics, the quarterback of the football team, the president elect of the Outing Club, and an officer of the Dramatic Association prominent in the production of student plays. The lowest thirteen men were decidedly without prominence in college activities.

To be sure, a certain amount of the difference between the two groups is to be attributed to the factor of intelligence, and indeed this factor can never be completely isolated in any study of personal differences. But an examination of the intelligence ratings and the scholarship marks of the two groups here studied shows that they differ by far less in respect to intelligence than in respect to aggressiveness. If the selection had been of the highest thirteen men in the class on the basis of intelligence ratings, only four of our most aggressive men would have qualified, and one of them would have been included in a list of the thirteen least intelligent. As against this we have the fact that five of the least aggressive were also rated among the thirteen least intelligent, and none of them among the thirteen most intelligent. The average of the scholarship marks of the thirteen most aggressive men is 2.5 on a scale of 4.0 as the maximum. This is about .3 of a point higher than the general average of the college. The average of the thirteen least aggressive men is 1.7, about .5 of a point lower than the college average. There is then a real difference in intelligence in favor of the most aggressive group, almost necessarily so, but it is not to be compared with the difference in aggressiveness. Not one of the highest group had an average rating of less than B on aggressiveness, and not one of the lowest group had an average rating of more than D on the same trait.

Having thus found two groups who differed in nothing so much as in aggressiveness, the next step was to try our tests which seemed calculated to give marked and consistent differences of results for the most and the least aggressive groups. The assumption was that such tests would have considerable significance as affording a measure of the trait. It was hardly to be hoped that any single test would show a point for point correspondence with the existence of the trait,—though the eye movement test does almost approximate such a correspondence. But the expectation was that a grouping of several significant tests would offer a highly suggestive picture of the trait in a given individual,—an aggressiveness profile,—to use Downey's expression.

TABLE I  
EYE MOVEMENT TEST

	Name	No. of Eye Movements	
Most Aggressive.....	C.N.	0	Total 6
	Z.J.	0	
	A.C.	2	
	R.M.	1	
	N.A.	0	
	J.G.	1	
	E.B.	0	
	B.W.	0	
	R.W.	1	
	H.C.	0	
	A.J.	0	
	E.L.	1	
	D.R.	0	
Least Aggressive.....	H.M.	7	Total 72
	J.J.	1	
	E.R.	4	
	W.P.	4	
	N.F.	4	
	P.G.	5	
	F.B.	7	
	J.N.	0	
	G.D.	5	
	R.B.	16	
	S.H.	0	
	R.T.	5	
	B.R.	14	

## I

## EYE CONTROL IN PERSONAL INTERVIEW

Common sense has it that the shifty eye is generally a sure sign of personal weakness, if not of downright dishonesty. The first test was designed with a view to bringing this element of behavior into quantitative relation with the trait aggressiveness. Each of the twenty-six subjects was required to perform a somewhat difficult series of mental additions while constantly returning the fixed gaze of the instructor who sat facing him. The addition series were standardized as to difficulty as described in test II, and the subject performed

them with the knowledge that accuracy and speed were essential. But he was emphatically instructed that under no circumstances should he let his gaze wander from that of the man facing him, as all movements of the eyes were to affect his score seriously.

The details of the addition series are described in the following test. The record for each subject is the total number of eye movements recorded against him during five series of additions. Table I gives the records of the twenty-six subjects.

It will be seen that a total of 72 movements are recorded against the least aggressive as compared with 6 against the most aggressive men. Not one of the aggressive group averted his gaze more than twice during the five series of additions, whereas ten of the thirteen least aggressive subjects shifted their eyes four times or more. If a steady eye does not positively guarantee the presence of aggressiveness, as may be judged by the fact that three of the least aggressive subjects were able to maintain practically constant fixation, a marked lack in this respect is almost invariably accompanied by a lack of aggressiveness.

The question naturally arises whether eye control in personal interview is not perhaps more a matter of "reliability" than of aggressiveness. To test this point the thirteen most reliable and the thirteen least reliable men in the class, as determined by faculty ratings, were submitted to the same experiment. Eight of the thirteen "least reliable" were able to keep within a limit of two eye movements, and the total of the thirteen subjects was 41 movements as compared with 72 for the thirteen least aggressive. Six men were common to the two groups, which would indicate a fairly high correlation between the two traits, but the record for eye movements evidently points to a much more definite relation between eye control and aggressiveness than is the case with reliability.

Whether we compare the average number of movements of the aggressive and the unaggressive subjects, or compare the number of subjects in the two groups who lost control to the extent of making two or more eye movements, we find evidence of at least ten times as much control in the upper group as in the lower. Thus the simple behavioristic fact of the ability to look another person in the eye seems to have such a high significance regarding the presence or absence of aggressiveness as to warrant giving it an extremely prominent place in any scoring method devised as a measure of this trait. The correspondence is in fact so close as to justify the generalization that a stop watch and a pair of fixed eyes are the

only indispensable laboratory equipment necessary for estimating roughly the degree of aggressiveness in at least four-fifths of the subjects.

## II

### FEAR DISTRACTION TESTS

The purpose of the distraction tests was to determine how far the ability to resist a type of distraction involving a certain element of fear is related to the possession or lack of aggressiveness. Preliminary to all of these tests each subject was required to practice fifty times performing mental additions according to a uniform plan involving constant increments of one, through a series of nine additions. If such a series began with 36, it would continue; 37, 39, 42, 46, 51, 57, 64, 72, 81. If it began at 29, it would continue; 30, 32, 35, 39, 44, 50, 57, 65, 74. Thus a series might begin with any number and would consist of nine additions, the numbers added being: 1, 2, 3, 4, 5, 6, 7, 8, and 9. On the day prior to the experiment each subject had practiced such additions with fifty series. This amount of practice seemed to bring each man reasonably near to the limit of his natural ability at such a task. The thirteen most aggressive subjects were able after this amount of practice to perform ten series of additions under normal conditions at an average time of 12.6 seconds. The thirteen least aggressive subjects averaged 15.6 seconds for the same ten series under normal conditions. The median of the thirteen mean variations for the normal series of the aggressive subjects was 2.0 seconds; that for the least aggressive subjects was 2.3 seconds. The difference in the normal variability of the two groups was thus negligibly small. In the comparison of any two individuals the factor of normal variability in the addition series will have to be taken into account, but for purposes of comparing these two groups as a whole it is of less importance.

### DISTRACTION BY STARING

Immediately after the ten normal series were completed the subjects were asked to perform the same task under conditions of slight emotional distraction. The first of these distractions was that of having to meet the fixed gaze of the instructor facing the subject, as described in the eye movement test. This distraction record is in fact the time record which was obtained simultaneously with the eye movement score. While one experimenter counted the eye movements which gave the



record for Table I, another timed the five series of additions performed under these conditions. The average increase of time as compared with the normal average is obtained as the measure of the subject's distraction. Table II gives the results for this test.

TABLE II  
DISTRACTION TEST—STARING

	Name	Time Increase With Test	A.D. Without Distraction	No. of Cases in which Distraction Time is Greater Than A.D.	
Most Aggressive	C.N.	-1.9	.6	...	Total 3
	Z.J.	.4	2.3	...	
	A.C.	.4	2.8	...	
	R.M.	-.1	2.0	...	
	N.A.	3.0	1.3	1	
	J.G.	-.4	2.5	...	
	E.B.	5.1	1.2	1	
	B.W.	-.1	3.5	...	
	R.W.	0.	.8	...	
	H.C.	.5	2.8	...	
	A.J.	.9	2.5	...	
	E.L.	2.9	1.0	1	
	D.R.	-1.4	1.3	...	
Least Aggressive	H.M.	9.7	2.5	1	Total 8
	J.J.	1.4	3.5	...	
	E.R.	5.3	2.7	1	
	W.P.	2.5	2.1	1	
	N.F.	-.5	2.3	...	
	P.G.	-.6	1.8	...	
	F.B.	4.9	2.4	1	
	J.N.	-2.0	2.6	...	
	C.D.	-.3	1.5	...	
	R.B.	2.8	1.4	1	
	S.H.	4.8	1.5	1	
	R.T.	5.0	4.2	1	
	B.R.	8.9	2.0	1	

The most aggressive subjects are listed in the upper half of the table, and the least aggressive subjects in the lower half. The first column of numbers shows the average amount of time increase of each subject under the distraction conditions, as measured by comparing the average time for the five staring tests with the average time for the ten normal trials. When

the time made in spite of distraction was more rapid than normal, this is indicated by a minus sign before the number. The middle column gives the average deviations for each subject in the normal series, and the right hand column indicates the cases in which the amount of time increase under distracting conditions was greater than the average deviation under normal conditions. The results may be summarized as follows: for the aggressive subjects the staring caused an average delay of .4 of a second; for the unaggressive subjects it caused an average delay of 3.2 seconds. Only three of the aggressive subjects were delayed by an amount equal to their average deviations under normal conditions, whereas eight of the unaggressive subjects have a distraction time greater than the A. D. The results indicate therefore approximately three times as much probability that an unaggressive person when stared at will be deterred to a considerable degree than that an aggressive person will be thus delayed.

#### DISTRACTION BY ELECTRIC SHOCK

In the second of the three distraction tests the element of distraction was the expectation of an electric shock which was to come during or at the end of each of the five series of additions. The subject was told that he would receive shocks, which might be of any intensity from 75 to 220 volts, and which might be expected at any time during or between the addition series. Actually he was not given more than 75 volts at any time, and the shock was always given just as he was finishing his last number in a series, so that the distraction was almost wholly imaginary rather than sensory.

The accompanying picture shows the apparatus used in the shock test.

The base of the apparatus consists of a 9 by 18 inch board, divided into two square sections by a board upright 5 inches high. An electric light plug and five sockets, fitted with 60 watt lamps, are on one side of the partition. The sockets are interconnected by a combination of series and parallel wiring. On the other side of the partition are a small voltmeter and two binding posts. The voltmeter and the posts are connected to the lamp sockets by four wires running through the partition. Two brass hand electrodes are connected one to each binding post. By tightening or loosening certain lamp bulbs a difference of potential of 75 or 220 volts can be sent either through the meter, or to the electrodes, or to both. The lamps, of course, are lighted whenever they are thrown in the circuit,

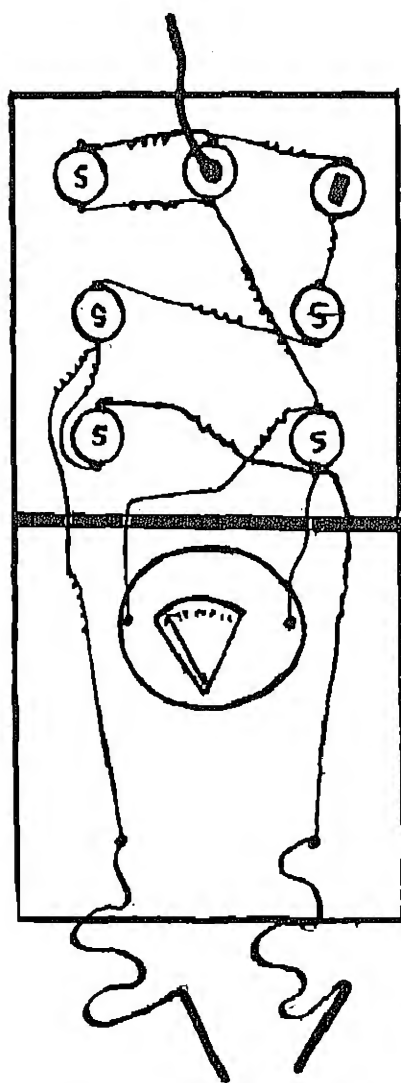
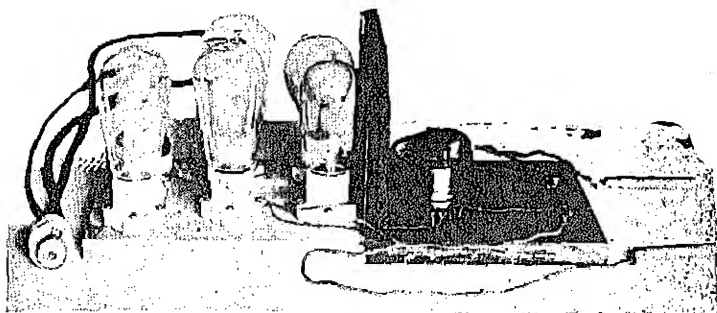


Diagram Showing Wiring for the Electric Shock Apparatus.



Electric Shock Apparatus

but they are partially shielded by the partition, and carbon lamps are used so as not to produce a bright distracting light.

The subject is seated by a table on which the apparatus rests. The experimenter hands the subject the electrodes and says: "This is an electric apparatus. As you begin your additions I shall begin manipulating these light bulbs. When I loosen or tighten certain bulbs you will receive an electric shock. It may be weak or strong. It is not dangerous, but may be very uncomfortable. Just to show how it feels I will give you a weak shock. Watch the pointer on the voltmeter." The subject is then given a shock of 75 volts. The electrodes are laid on the table, and the subject is asked to watch the voltmeter while another manipulation sends 220 volts through it. He is again told to take the electrodes in hand, and the test begins. The experimenter continually manipulates the bulbs during the adding, and just at the last addition in each series the subject receives a shock of 75 volts. No shock is given during the adding, but the subject is continually in a state of expectancy. Each subject is bound by a promise of secrecy regarding everything about the experiment. The tests were completed in two days and no subject had any knowledge further than that there was to be some kind of electric shock.

The results are given in Table III, and according to the tabular scheme already described for Table II.

The figures for time increases in the left hand column show that five of the aggressive subjects resisted the distraction with complete success, whereas all of the twelve, unaggressive subjects were positively affected to the extent of an increase of 2.5 seconds or more. The average shock-delay of the least aggressive subjects was 6.1 seconds as compared with

TABLE III  
DISTRACTION TEST—SHOCK

	Name	Time Increase With Test	A.D. Without Distraction	No. of Cases in which Distraction Time is Greater Than A.D.	
Most Aggressive	C.N.	2.9	.6	1	
	Z.J.	2.5	2.3	1	
	A.C.	10.0	2.8	1	
	R.M.	3.0	2.0	1	
	N.A.	4.8	1.3	1	
	J.G.	5.8	2.5	1	
	E.B.	3.9	1.2	1	
	B.W.	-2.3	3.5	...	
	R.W.	.4	.8	...	
	H.C.	0.	2.8	...	
	A.J.	-.9	2.5	...	
	E.L.	8.0	1.0	1	
	D.R.	-.4	1.3	...	
					Total 8
Least Aggressive	H.M.	10.1	2.5	1	
	J.J.	9.9	3.5	1	
	E.R.	6.0	2.7	1	
	W.P.	17.1	2.1	1	
	N.F.	3.3	2.3	1	
	P.G.	...	1.8	...	
	F.B.	2.5	2.4	1	
	J.N.	2.6	2.6	1	
	G.D.	4.7	1.8	1	
	R.B.	5.5	1.4	1	
	S.H.	3.8	1.5	1	
	R.T.	2.9	4.2	...	
	B.R.	4.7	2.0	1	
					Total 11 (One score missing)

an average of 2.2 seconds for the aggressive group. The individual variations among the aggressive subjects seem to indicate that this test has a less specific relation to aggressiveness than does the staring test, and this is somewhat borne out by the fact that when the same test was given to the thirteen most reliable and the thirteen least reliable men in the class, the difference was by far more striking than for the aggressiveness groups. Only one of the thirteen "most reliable" men had a shock-delay of more than three seconds, whereas eleven of the thirteen "least reliable" men had delays upwards of three seconds. However, the fact that the ability to with-

stand shock distraction is possessed by 40 per cent of the aggressive subjects and by none of the unaggressive ones justifies the inclusion of this test as having considerable symptomatic value regarding the trait.

#### DISTRACTION BY SNAKE

In the third test the method of distraction was to place a dead snake, suitably coiled, and pinned to a cork board, about ten inches in front of the face of the subject during one-half of his adding. Only one series of additions was recorded for this test, as it was found that adaptation was very rapid. This means, of course, that the results have a larger probable error than those recorded for the other types of distraction. The detailed record is given in Table IV.

TABLE IV  
DISTRACTION TEST—SNAKE

	Name	Time Increase With Test	A.D. Without Distraction	No. of Cases in which Distraction Time is Greater Than A.D.	
Most Aggressive	C.N.	9.1	.6	1	
	Z.J.	4.4	2.3	1	
	A.C.	6.6	2.8	1	
	R.M.	2.7	2.0	1	
	N.A.	3.6	1.3	1	
	J.G.	10.6	2.5	1	
	E.B.	7.7	1.2	1	
	B.W.	4.3	3.5	1	
	R.W.	1.4	.8	1	
	H.C.	— .5	2.8	...	
	A.J.	4.3	2.5	1	
	E.L.	5.9	1.0	1	
	D.R.	0.	1.3	...	
				Total 11	
Least Aggressive	H.M.	9.7	2.5	1	
	J.J.	11.4	3.3	1	
	E.R.	12.6	2.7	1	
	W.P.	.1	2.1	...	
	N.F.	1.1	2.3	...	
	P.G.	9.8	1.8	1	
	F.B.	21.7	2.4	1	
	J.N.	.8	2.6	...	
	G.D.	8.7	1.8	1	
	R.B.	11.6	1.4	1	
	S.H.	3.2	1.5	1	
	R.T.	7.0	4.2	1	
	B.R.	9.5	2.0	1	
				Total 10	

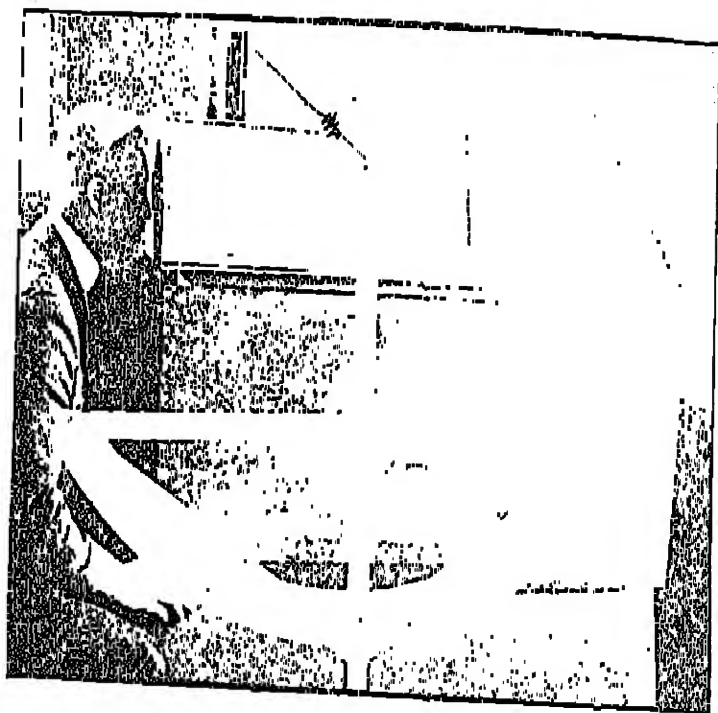
The time increase for practically all of the subjects was greater than the average deviation in the normal series, but the average delay of 8.2 seconds for the unaggressive subjects was practically double the 4.6 seconds for the aggressive ones. The most noticeable fact here was the greater liability of the unaggressive subjects to become somewhat panicky. Generally a delay of nine seconds or more was accompanied by very definite movements of withdrawal or avoidance, and it will be seen from the figures in the first column of the table that eight of the unaggressive men were delayed by nine seconds or more, as compared with only two of the aggressive men who were so delayed. The unaggressive man's liability to panic in the presence of a snake seems to be about four times as great as that of the aggressive man. It is interesting to note further that this test showed no relation to reliability.

### III

#### WORD ASSOCIATION TESTS

The word association tests were designed with a view to finding out whether the study of responses to selected stimulus words would give reliable data for the differentiation of men as aggressive or unaggressive. The six words selected as stimuli for this test were: 'enterprise,' 'success,' 'danger,' 'death,' 'opponent,' and 'company.' 'Enterprise' and 'success' seemed likely to contain positive and definite suggestions for the aggressive man, and 'death' and 'danger' to be more full of definite suggestions to the weaker subjects. 'Opponent' offered the concrete situation of personal contest, and 'company,' being adapted to so many possible meanings, seemed suited to the purpose of uncovering the readier tendencies in the mind of the subject.

**Apparatus.** The use of a lip key or similar recording device in circuit with a chronoscope was soon found unsatisfactory on account of the frequency with which the subjects tended to make lip movements, or even to begin to vocalize, before they were actually ready with a response. Whenever this occurred, it meant not only that the time record was spoiled, but that the stimulus word could never be given to that subject again under the original conditions. Moreover, oral presentation of words proved somewhat unsatisfactory on account of the possibility of slight misunderstandings, and the slight variations in emotional suggestion on the part of the experimenter. The apparatus finally constructed is shown in the accompanying picture.



Word Exposure and Association Time Apparatus



It consists of two main parts, the timing apparatus and the exposure box. The first of these consists of a standard, a pendulum, and a graduated metal arc. The standard is 42 inches high and supported by a base. The pendulum is a metal rod, 40 inches long, and is attached 10 inches from one end to the top of the standard. The shorter end of the pendulum is weighted so that it will make a half swing in exactly one second. The metal arc is one-fifth the length of the circumference of the pendulum swing. It is attached to the standard, one half on either side, just back of the path of the end of the pendulum. The arc is graduated so that time may be measured to within 5 hundredths of a second from the pendulum swing.

The exposure box is 10x12x20 inches and is closed except for a 2x4 inch opening at one end and a 1x2 inch opening at the other. From a revolving spool, attached just above the smaller opening, a sheet of paper 2 inches wide and 70 inches long extends to another similar spool just under the opening. The paper may be wound from one spool to another by means of handles attached to the spools. A piece of metal fits over the paper in front of the opening of the box so that it excludes all light. The stimulus words are arranged about an inch apart on the strip of paper, and by proper adjustment of the spools any word may be made ready for the exposure which is to result from the closing of the light circuit. For this purpose two 6 candle power lights are provided, one at either side of the larger opening of the box. These are wired in series with a battery and connected with the catch which suspends the pendulum in such a way that the moving of the catch to release the pendulum closes the light circuit, and thus exposes the stimulus word to view.

The subject sits with his eyes close to the larger opening of the exposure box, looking towards the paper about to be exposed at the other end of the box when the light connection is made. When the lights are not on, the box is dark, and the word on the paper is therefore not visible. The subject is given instructions to look for the word that is to be exposed, and to respond as quickly as he can with the first word that he can associate with it. A second experimenter watches the pendulum, and by noting the position of the pendulum on the graduated arc at the moment of the response, he is able to record the time to within 5 hundredths of a second.

The two sources of error in determining word reaction time by this method are: (1) the auditory reaction time of the second experimenter and (2) the errors in reading the swing

of the pendulum from the graduations of the arc. The first of these is probably very nearly constant for the same experimenter throughout a series of observations. The latter error may be in either direction but is seldom more than 5 hundredths of a second. In general the time records here presented are a small fraction of a second too long. But as this tendency is constant in all the records, and the method offers the subject no possible way of stealing a march on the experimenter, it seems highly satisfactory.

The results of the association experiment are set forth in Table V.

The six stimulus words are indicated in the horizontal line at the top of the table. Each subject's responses to the six words are given after his initials. Just above each of the response words is a large number on the right which indicates the time record in hundredths of a second. The small number just above the response word shows the frequency with which that response was given when the experiment was repeated with 78 subjects. Twenty-two of the original twenty-six subjects were available for this experiment. The upper part of the table contains the results for the eleven very aggressive subjects, and the lower half the results for the eleven least aggressive ones.

There is evidently a significant difference between the two groups in the character of the words chosen. In response to the words 'enterprise' and 'success' the aggressive subjects are almost always definite and positive. Of their twenty-one responses fourteen are decidedly colorful;—'initiative,' 'push,' 'money' twice, 'activity,' 'scheme,' 'undertake,' 'ambition,' 'power,' 'gain,' 'win' twice, 'wealth' and 'advance.' Five responses were sufficiently general to be classified as of doubtful significance;—'building,' and 'business' four times. One was apparently based on verbal similarity;—'success'—'successful.' Only one response;—'failure' was negative. The twenty-two responses of the eleven least aggressive men to the same two words were strikingly different. Only three were positive and colorful;—'work' and 'ambition,' the latter occurring twice. Twelve are doubtful, colorless, or passive;—'physics,' 'American Magazine,' 'oil,' 'friends,' 'good,' 'college,' 'happiness,' 'business' four times, and 'industry.' The negative response;—'failure' occurred four times. Briefly, there is four times as much probability of a definite, forward-looking response to 'enterprise' and 'success' from the aggressive man as from the unaggressive. Furthermore, there is

TABLE V  
MOST AGGRESSIVE SUBJECTS

Stimulus Words						
	Company	Enterprise	Opponent	Success	Danger	Death
Name C.N.	5 400 army	1 140 initiative	1 260 box	1 200 power	34 150 fear	1 275 cotton
Z.J.	1 220 battery	2 175 building	26 135 enemy			1 135 living
A.C.	3 200 girl	22 105 business	1 255 politics	2 170 money	2 95 dread	33 175 life
R.M.	1 190 friendly	1 190 push	2 400 hatred	2 135 gain	34 100 fear	33 130 life
N.A.	3 200 girl	2 140 money	2 175 opposite	3 200 win	34 135 fear	3 140 die
J.G.	2 130 people	1 130 activity		1 135 wealth	34 135 fear	11 230 fear
E.B.	2 200 alone	1 310 scheme	26 200 enemy	3 210 win	2 150 dread	33 115 life
R.W.	1 130 association	2 100 undertake	7 95 ice	1 200 successful	34 95 fear	2 135 death
H.C.	2 210 business	22 95 business	1 300 war	1 250 advance	34 175 fear	11 340 fear
A.J.	1 130 insurance	22 130 business	26 195 enemy	5 140 business	1 200 car	33 155 life
E.L.	2 135 society	4 180 ambition	1 130 adversary	25 135 failure	8 150 safety	133 100 life

LEAST AGGRESSIVE SUBJECTS

H.M.	1 145 good	22 145 business	26 155 enemy	1 420 Annet. Mag.	2 210 red flag	2 155 grave
J.J.	2 150 people	1 160 enterprise	26 160 enemy	2 130 happiness	8 200 safety	33 140 life
E.R.	2 120 group	4 140 ambition	9 160 fight	2 200 succeed	2 400 death	11 140 fear
W.P.	3 100 girl	1 280 oil	2 85 football	1 140 college	1 135 money	1 115 killed
N.F.	2 115 alone	1 195 encounter	2 100 antagonist	5 300 business	2 430 act	2 130 birth
F.B.	2 200 man	22 235 business	3 200 man	2 240 good	2 150 red flag	1 150 man
J.N.	5 200 army	1 400 work	26 200 enemy	25 95 failure	1 320 enemy	33 160 life
G.D.	9 145 friend	3 135 industry	9 150 friend	25 150 failure	1 300 fly	33 100 life
R.B.	6 230 friends	22 230 business	26 140 enemy	1 180 friends	34 95 fear	3 285 funeral
S.H.	6 105 friends	1 250 failure	8 130 friend	25 115 failure	34 145 fear	33 120 life
R.T.	1 400 Ford	4 330 ambition	1 300 hardware	1 240 physics	34 160 fear	1 190 accident

only one fourth as much probability of his giving a negative response, such as 'failure.'

Almost as clear cut is the difference in the character of the responses of the two groups to the words 'danger' and 'death.' In this case the aggressive responses are mostly very general and commonplace, while those of the unaggressive are more concrete and vivid. Of the twenty-one responses of the aggressive subjects only three seem to indicate anything resembling a vivid, personal experience. These words were 'car' and 'dread,' the latter occurring twice. The response 'cotton' to the stimulus word 'death' is doubtful, and the remaining seventeen are either so common as to be meaningless, or else are mere verbal variations of the stimulus word:—'life' five times, 'fear' seven times, 'safety,' 'die,' 'death.' By contrast, the unaggressive subjects present at least 10 out of 22 responses which are colorful and suggestive of personal interpretation:—'money,' 'fly,' 'act,' 'enemy,' 'red flag' twice, 'grave,' 'killed,' 'funeral,' 'accident.' The remaining twelve are either doubtful or merely verbal:—'life' four times, 'fear' four times, 'safety,' 'birth,' 'man,' and 'death.' In short, the probability of a definite and vivid response to the words 'danger' and 'death' is four times as great for the unaggressive man as for the aggressive man.

The responses to 'company' show exactly twice as many military, sexual, and commercial associations for the aggressive as for the unaggressive subjects. 'Army' and 'battery' in the aggressive list are to be compared with 'army' once in the lower list; 'girl' occurs twice in the upper list to once in the lower list; and 'business' and 'insurance' in the upper list are to be compared with 'Ford' in the lower list. The evidence, which is slight, points to the conclusion that aggressiveness doubles the probability of a definite and energetic response to the word 'company.'

The responses to the word 'opponent' are extremely difficult to interpret, and seem to show such an influence of the language habit that the word is probably not a suitable one for giving scorable results. The aggressive responses:—'box,' 'politics,' 'hatred,' and possibly 'war' are more pointed expressions of personal antagonism than any of the words in the unaggressive list except 'fight;' but 'football' and 'hardware' leave considerable doubt, and both lists contain a high per cent of merely verbal associations.

Confining our statement to the five words which afforded the most definite responses, we may say that the very aggressive man is four times as likely to be positive

and definite in his responses to 'enterprise' and 'success,' twice as likely to give an energetic type of response to 'company,' and only one-fourth as likely to respond definitely and vividly to 'danger' and 'death' as the very unaggressive man. These differences are large enough to warrant the use of the word responses as one of the minor scoring elements in a measurement of aggressiveness.

The time records for the responses show a marked difference for the stimulus words 'danger' and 'enterprise.' Otherwise the average differences for the aggressive and the unaggressive groups are less than half a second. For 'danger' the unaggressive subjects take an average time of 2.31 seconds as compared with an average time of 1.26 seconds for the aggressive subjects. For 'enterprise' the unaggressive subjects average 2.55 seconds and the aggressive subjects 1.54 seconds. Failures to respond are not here reckoned in averaging the time, for the reason that such failures were due to imperfect exposure of a few words in the early part of the experiment.

Seven of the unaggressive responses to 'enterprise' required two seconds or more, as compared with one such case among the aggressive subjects. Likewise six of the unaggressive subjects required two seconds or more to respond to the word 'danger,' as compared with one such case among the aggressive subjects. This difference is not due to the greater average quickness of response of the aggressive subjects, as their averages for the remaining four words are actually about a fifth of a second slower than the corresponding average for the unaggressive subjects. Apparently it is due to some peculiar relation of the two words to the two classes of subjects. The obvious assumption in regard to 'danger' is that it comes near to awakening true emotional disturbances, much more nearly so in the unaggressive subjects. Their slowness in responding to 'enterprise' might be due either to emotional inhibition or to paucity of ideational material, either a matter of self reproach or of an unfamiliar concept. Whatever the theoretical explanation, the time record of the two groups of subjects is large enough and constant enough to justify its use in differentiating the two types of individuals.

#### CONCLUSION

We are now ready to propose a method of scoring aggressiveness on the basis of our tests. In order to have a perfect record on the trait, a man should be able to maintain perfect control of his eyes in the interview test; he should withstand

with complete success all the three types of distraction; and in the association tests he should respond definitely and positively to 'enterprise,' 'success,' and 'company,' and indefinitely to 'danger,' and 'death,' moreover, his reaction time to 'enterprise' and 'danger' should not be noticeably longer than his normal word reaction time. The scoring method proposed makes deductions from 100 points for each failure of a subject to meet satisfactorily the four types of requirement just mentioned. Corresponding to the relative importance of the different tests for this purpose a maximum possible deduction of 50 points is allowed for eye movements in test I, a maximum of 25 points each for the distraction and association tests. Stated more in detail, the scheme of deductions for each type of failure is as follows: In the interview test 5 points are deducted for every movement up to ten movements. The maximum deduction is therefore 50 points. For the staring distraction record two points are deducted for each second of time increase in excess of three seconds above his normal adding time. The maximum deduction allowed is 10 points. For the shock distraction 2 points are deducted for every second of time increase in excess of five seconds. The maximum deduction allowed is 10 points. For the snake distraction test 1 point is deducted for each second of time increase in excess of seven seconds. The maximum deduction allowed is 5 points. In the word association test 2 points are deducted for each negative response to 'enterprise,' 'success,' or 'company' and for each vivid, personal type of response to 'death' or 'danger.' One point is deducted for each colorless or doubtful response to 'enterprise,' 'success' or 'company.' The maximum deduction on the basis of the word content of responses is thus 10 points. For the time of response to 'enterprise' and 'danger' 1 point is deducted for every fifth of a second in excess of two seconds. The maximum deduction allowed on this account is 15 points.

The score of any individual subject in aggressiveness is to be obtained by totaling his deductions and subtracting the amount from 100 as shown in table VI. On this basis the scores of our thirteen most aggressive subjects are as follows: C. N. 97, Z. J. 100, A. C. 79, R. M. 94, N. A. 98, J. G. 86, E. B. 83, B. W. 100 (lacking word test), R. W. 93, H. C. 98, A. J. 97, E. L. 84, D. R. 100 (lacking the word test). The scores for the thirteen least aggressive subjects are as follows: H. M. 35, J. J. 79, E. R. 55, W. P. 60, N. F. 59, P. G. 72 (lacking word test), F. B. 36, J. M. 81, G. D. 63, R. B. 37, S. H. 88, R. T. 61, B. R. 38 (lacking word test).

TABLE VI  
FINAL SCORES OF 26 SUBJECTS  
Most Aggressive

	Eye Movement	Staring	Shock	Snake	Words	Time	Score
C.N.	..	..	..	2	1	..	97
Z.J.	..	..	..	.	1	..	100
A.C.	10	..	10	.	1	..	79
R.M.	5	..	..	.	1	..	94
N.A.	..	2	..	.	..	..	98
J.G.	5	..	2	4	3	..	86
E.B.	..	6	..	1	4	6	83
B.W.	..	..	..	.	x	x	100
R.W.	5	..	..	.	2	..	93
H.C.	..	..	..	.	2	..	98
A.J.	..	..	..	.	3	..	97
E.L.	5	2	6	.	3	..	84
D.R.	..	..	..	.	x	x	100
			Least	Aggressive			
H.M.	35	10	10	3	6	1	35
J.J.	5	..	10	4	2	..	79
E.R.	20	4	2	5	4	10	55
W.P.	20	..	10	.	6	4	60
N.F.	20	..	..	.	6	15	59
P.G.	25	..	..	3	x	x	72
F.B.	35	4	..	5	5	15	36
J.N.	..	..	..	.	4	15	81
G.D.	25	..	..	2	5	5	63
R.B.	50	..	2	5	4	2	37
S.H.	..	4	..	.	5	3	88
R.T.	25	4	..	.	3	7	61
B.R.	50	10	..	2	x	x	38

The average score of the aggressive group is 93 and that of the unaggressive, is 58.8. None of the aggressive subjects score less than 79, whereas nine of the least aggressive make less than 64, and one of the others is incomplete. Three of the twenty-six subjects, J. N., S. H., and J. J. appear as noticeable exceptions. It is the opinion of the writers, based on considerable acquaintance with the subjects, that the error in regard to at least two of them lies in original personal ratings rather than in the finding of the test. In any event we seem to be justified in stating that there is hardly one chance in twenty-five that a man weak in aggressiveness would score as high as 85; and there is almost no chance that a very aggressive person will score lower than 70. The writers believe that this test approximates a true measurement of aggressiveness more nearly than does the Army Alpha examination approximate the measurement of intelligence.

Moreover, the test can be given in modified and abbreviated form with no further equipment than a stop watch and a set of addition tables, with the additions all made according to the specifications in test II. In this case it would be necessary to omit the shock and snake distraction tests; also to give the words for the association test orally, and to time the responses with a stop watch. The loss of two of the 10 point tests can be compensated for in the scoring by adding 25% to the total of the deductions made on the basis of the remaining tests.

Aside from the time required of each individual subject for practicing his 50 additions with the tables, the actual time of administering the abbreviated form if the test should not be more than three minute. The longer form with apparatus requires about seven minutes per subject.



# THE BEDAUX-UNIT PRINCIPLE OF INDUSTRIAL MEASUREMENT

## ITS ORIGIN, ITS LAWS AND APPLICATIONS

---

By CHARLES E. BEDAUX

---

(This article is copyrighted by the author, and reproduction in any form is forbidden except by his express permission.)

Ever since industry reached the tremendous proportions afforded by the mechanical era, the need has been felt for a system of measurements permitting the exact determination of the quantitative value of the human physical effort.

It is due to the lack of such a system that to-day we meet deeply set differences between management and labor. Either side ignoring the value of a given effort made it possible for the employee to practice, undetected, a restraint in production, while the employer, in some instances, protected by the same ignorance, has been able to arbitrarily decrease the money value of the operation performed.

A correct system of measurement of the industrial effort has many applications:

1. It permits a protective and adequate system of remuneration;

2. It gauges the capacity of the human being or of a group of employees comprising one department, permits an exact comparison between departments, between plants, and automatically indicates temporary or permanent conditions that are conducive to failure;

3. It may be used as a basis for apportionment of the overhead burden;

4. It allows an automatic and constant comparison between the standard and actual labor cost;

5. It renders scientific planning and scheduling of manufacture possible;

6. It is the only foundation on which any form of labor representation in shop management may stand.

The exponents of the piece-work system have recognized that the basic weakness of their method lies in the inability to guarantee the rate set, as it is based on the monetary value of labor, a commodity which fluctuates. To guarantee the

rate is to expose the manufacturer to failure, in case of a general decrease in the labor market. To cut the rate destroys the confidence of the working man and compels a restraint in production.

The monetary value of labor, being a variable, cannot be used as a means to measure industrial effort.

Because of fluctuations in this value, not only between periods of depression or growth, but also within each period, among the various classes of labor, it has been impossible for the manufacturer to gauge the respective productive capacities of his various departments, or of his entire plant, in comparison with a competitive or co-operative institution.

The result has been that many plants, sincerely believing that their productivity rate was high, have failed because of their error, the detection of which was impossible without an adequate system of measurement of the human effort applied to industry.

It is also because of the absence of such a system that the monetary value of productive labor has been recorded for cost purposes by the taking of the time of the beginning and the ending of each operation. This performance is generally inaccurate, or, when accurate, necessitates waiting time for recording—a waste of labor, both in the productive and in the clerical forces.

To that source also must be traced the erroneous distribution, too often met, of the overhead burden, wherein an operation costing one dollar (\$1.00) in productive labor, performed on a machine worth one thousand dollars (\$1,000.00) and depreciating in twenty years, for instance, will be charged with the same amount of overhead as an operation also costing one dollar (\$1.00) in productive labor, but performed on a machine the cost of which is one hundred thousands dollars (\$100,000.00) with a rate of depreciation that must be absorbed in ten years.

Because of the lack of an adequate system of measurement, incapable operators have been kept in positions that have been refused to men who were best indicated to hold them; promotion has been arbitrary and illogical, thus entailing losses to both capital and labor and creating a state of mind on either side which is wholly destructive.

### *Part Two*

For many years we have seen the need of a unit of measure and our study first led us to determine that such a system could not be based on monetary values which fluctuate. We

## FAN MACHINE ASSEMBLY

then established that industrial effort was surrounded by two constants, i.e., man and time.

While the productive value changes with every human being, the average power of a large group, taken in different generations, remains constant, the variations being caused by difference in facilities and not in men.

Time is, of course, also a constant. It therefore appeared, on first consideration, that time could be used as a unit of measurement for industrial effort.

We experimented with the use of these two factors and found that the function we call "industrial work" is composed of three elements, i.e., effort, rest and unavoidable delays. We first established a system based on the element of effort alone and found it impractical, because erroneous. It brought us to determine that the percentage of rest to effort varies according to the nature of the operation.

Further studies brought us to formulate these two laws:

1. *For a muscular effort of a given power, the duration of work and rest periods is inversely proportional to the rapidity of the motion, and*

2. *For a muscular effort of a given power, the percentage of rest is directly proportional to the rapidity of the motion and completion of the cycle.*

By the first law it is meant that the more rapid the motion and the shorter the cycle, the more often the rest periods must be repeated. For instance, it has been found that marching in the army gives best results with fifty minutes' effort and ten minutes' rest, while in an operation requiring fast motion of the fingers, such as packing of very light packages, in one instance, the best results were obtained with routine three minutes' work periods and one minute's rest periods, although the percentage of muscle relaxation to muscle strain is actually greater than one to three.

The practice of setting definite rest and work periods to be observed by labor is not to be recommended except in rare cases. When they are considered it must be remembered that man, who as an average is a constant, as an individual offers a wide range of variation in power as well as in skill.

By the second law it is meant that the shorter the operation and more rapid its motion, because of light weight and scarcity of varied movements, the greater the percentage of rest must be.

We further found that a worker, performing a long operation, had a greater opportunity to increase his efficiency by combination of moves than is permitted to one having to

perform an operation the cycle of which is short, therefore demanding a positive routine of a few movements.

On the strength of the above conclusions, our next step led us to compile tables of all ratios of rest and unavoidable delays affecting the various forms of human physical effort in industry. We found it to vary from 15% of the actual productive time in long cycle hand assembly operations to 150% in mechanical operations, hand-fed, of a cycle of two seconds or less.

Experiments, analysis and practical experience born out of various applications of the above brought us to formulate a third law.

3. *In a large group of equally trained workers the average production of the tenth of the total number that is ablest does not exceed the double of the average production of the balance.*

The securing of the above information allowed us to establish the Bedaux Point System of Industrial Measurement. The unit was called a "Point." It has the following definition:

*A point is a fraction of a minute of work, plus a fraction of a minute of rest, the aggregate of which is always one minute, but the proportions of which vary according to the nature of the operation.*

An employee, man or woman, taking advantage of the normal amount of rest, produces therefore sixty points in one hour, this being the standard. An operator producing more than sixty points per hour is, at the end of the day, entitled to a reward based on the difference between the total points produced and the standard of sixty points per hour.

The operator, by application, can increase his production beyond sixty points per hour without fear of suffering an arbitrary decrease in his reward. The point value of his operation or operations, being based on time in relation to average human physical energy, two elements that are constant, is not, therefore, subject to change. To be logical, the manufacturer must guarantee the point value of any operation, as long as the said operation is not changed in process or material.

The exact knowledge of the number of points produced per hour by each man allows an intelligent system of promotion and the exact knowledge of the number of points produced per hour by a group of men or department permits an adequate remuneration of the leader or foreman. This protective system of remuneration based on the point is known as "The Point System of Graded Remuneration."

*Part Three*

The logical industrial enterprise is one which manufactures an article or articles in normal demand, made by normal men, performing under normal conditions.

The rest and delays allowances figuring in the establishment of a point representing normal conditions, the aggregate of the elements forming a point representing one minute, the logical manufacture must be profitable at a production of sixty points per hour.

Whenever operators produce more than sixty points per hour, the difference must figure in the element of reward of the operator and of those directly connected with production and whose co-operation labor seeks, with a view to secure the maximum production.

The gain of the manufacturer is realized in the reduction of the overhead burden per unit.

The industrial enterprise which finds any or all of its departments producing daily at a rate of less than sixty points per hour knows it is in danger and can therefore take immediate steps to bring the point hour to sixty or higher in any or all of its departments.

Manufacturers having branch plants can compare the point hour of each plant each day and thus measure the advance of the various units of their enterprise.

The aggregate point value of an article represents the standard labor cost. The accumulation of all the points charged to a given factory order allows automatic comparison of the actual cost against standard cost, this without the expensive recording of the time of beginning and ending of each operation.

The unit of measurement of the Bedaux system, the point, can be used in the distribution of overhead burden by determining in advance the cost of running one point through each manufacturing unit and multiplying this cost by the point value of all operations passing through, thus giving an exact overhead distribution.

Countless efforts have been made in the past to plan, schedule and dispatch material through manufacture. Many failures have resulted, for it must be remembered that planning is a relation of requirements to capacity.

To determine the time requirements of an article passed through different branches of manufacture, or to determine the time capacity of the various branches of manufacture through which the article is to be passed, one must have at

his disposal a unit of measure, which, in industrial effort, has heretofore been unknown.

With the Point, planning becomes an easy matter. The point value is known on every article in all its operations. The point capacity of all machines or branches of manufacture is known through the point hour. The work of planning thus becomes simply a routine detail relation of point value to point hour.

#### *Part Four*

Plans of profit-sharing have been established in the past with a questionable degree of success, only because there has been no way to measure the capacity of each individual when determining the share of the profit to which he was entitled. If the manufacturer sincerely desires to make his labor share in the profits, the amounts can be apportioned according to the point hour of each man in relation to his class of skill, which is gauged by his hourly rate.

Various forms of labor representation in shop management have lately been widely praised as the logical panacea for industrial ailments, but is it logical to organize a labor representation which is formed indifferently by men who are capable or incapable? Has it not always been the rule for man to master the capacity to perform before he assumes the capacity to advise? The point hour record of each man in his daily labor together with his mental capacity should alone determine his eligibility to representation of his fellow workers, whenever the manufacturer feels that he must adopt that form of management.

The most formidable obstacle to overcome in establishing industrial peace is the present restraint of hope caused by the industrial structure of to-day. The bench man remains a bench man. A press operator, an inspector, an automatic screw man, a lathe operator, all remain what they are. It has been so customary for industry to endeavor to keep a man wherever he is doing "fairly well" that labor itself has lost the thought of possibility of individual rise but in turn has reached a state of mind generally indifferent or else bitter and antagonistic.

Under the Point System of Industrial Measurement, an unskilled laborer who, by application, shows a point hour of, say, ninety, or thirty points above standard, this for a period of a year, is automatically entitled to a better position whenever there is a vacancy and he can begin such at a point hour of not less than sixty. If, by further application, he brings

himself, in this new and better position, to a point hour of ninety and maintains himself there, again he is automatically entitled to a still better position, at a higher hourly rate, providing he can begin his performance, through previous study, at a point hour of not less than sixty. This can be continued through all the branches to the very head, so that when a man is given the direction of a department because he has shown fitness in the elemental requirements of the executive position, it will be known that he has also proven himself capable in all the branches he commands.

At the present time the labor unions, knowing that the chances for advancement in productive labor are small because of industrial tradition, have endeavored to raise the scale of wages of the whole strata composed of the various classes of effort, a suicidal move, as if applied it only raises the price of the commodity purchased by the consuming class, which is largely composed of the class of people to which union labor belongs.

As illogical as this labor demand may seem, there is no logical answer to give other than to render possible the advancement of the individual through each class, according to the exact individual capacity, a capacity which to-day can be accurately measured by the Point without increasing but even decreasing the amount of clerical work usually found in most plants.



## A STUDY IN INDUSTRIAL PSYCHOLOGY—TESTS FOR SPECIAL ABILITIES

By ELSIE OSCIRIN BREGMAN

The following is the report of an investigation with the purpose of developing tests for the use of the Employment Department of R. H. Macy & Co., Inc., a large department store in New York City.

The problem resolved itself into a twofold one; first to determine whether or not the applicant who applies is of normal intelligence, and secondly, to determine, of the normal applicants, the specific kind of work to which they shall be appointed.

The first problem has been handled in a fairly adequate manner by an adapted form of the Trabue Completion Test. The results of this test will, it is hoped be presented very soon in a separate paper. This report concerns itself with the development of tests for specific abilities.

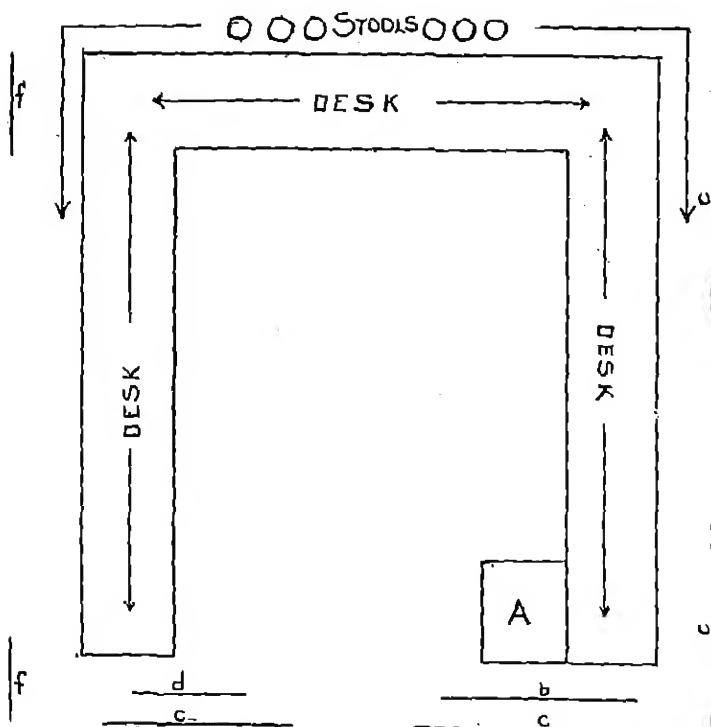
The greatest number of employees of a department store are salesclerks; the next in number are clerical workers. Consequently it was for tests that would indicate these two types of workers that investigation was first begun.

It has seemed worth while, in the light of the many new problems of administration which confront the investigator in industry, to add to the report of the findings of the investigation, an account of the manner in which the investigation itself was handled, since this phase is as vital to the success of the undertaking as an adequate knowledge of the scientific technique and procedure.

*Preliminaries.* In April of 1919 the writer came to the department store, whose employees number 5,000 to "experiment with vocational tests." Before beginning experimental work of any kind the writer went through a course of training, working in every department of the organization in every capacity—as wrapper, salesclerk, complaint clerk, etc., etc., through a miscellany of positions. This course enabled the writer not only to understand the organization in which she was to work, but to learn the personnel and personality of departments, to have more than a bowing acquaintance with department heads and executives, and, not the least important, to have a knowledge of useful kinds of records and of many sources of information which blueprints and verbal reviews did not give.

Immediately following this course of sprouts, the writer became an active member of the Employment Department, interviewed and appointed applicants and during three months (during which time an office which could be used as a laboratory was being built and equipped, and preliminaries to actual experimental testing, such as collection and printing of forms, etc., were carried on) acquired a grasp of the local labor situation, the sources of supply, the types, characteristics and social levels of applicants who presented themselves, and the ways in which appointments and judgments were made, etc., etc.

The following diagram illustrates the office laboratory which was built where the testing could be carried on with the minimum of noise and disturbance.



a. Examiner's Desk  
b. Bookcase

c. Windows  
d. Blackboard

e. Cupboard  
f. Doors

Before the testing itself took place, meetings of the executives, heads and assistant heads were called, at which the purpose, plans and method of the investigation were disclosed and explained as simply as possible, and cooperation was urged. In order to dispel prevalent bugaboos about the nature of mental tests, a five minute Trabue Completion Test was given.

*Obtaining the Criteria.* In view of the unfortunate waste that has not infrequently occurred when after extensive tests had been made, records of work and ability were unobtainable, it seemed wise to secure the criteria before testing, the more so that some coherent scheme of rating had to be worked out in an organization where for two thousand salesclerks there were as many as 200 department and section heads.

*Salesclerks.* Considering first the salesclerks, two methods of determining their ability were available (1) the actual weekly sales or production record made by each clerk, (2) the opinions of department heads, of which there are for each department

1. Buyer
2. Assistant Buyer
3. Section Manager or Floorman.

*Production Record.* Since the gross amount of sales varies for each department and from week to week, according to season and weather, it was not possible to throw all sales on to one scale for all time. The following method was used.

A simple frequency table was made showing the sales for each department for each week. From this the upper and lower quartile limits were determined, and record was made for each clerk, whether he sold within the interquartile range, above it, or below it. Such a record was kept for ten weeks and from this was determined whether a clerk sold characteristically within, above or below the average range.

*Ratings.* This information was supplemented by the ratings of department heads, which were obtained in the following manner:

The accompanying form was sent to each buyer and assistant buyer and section manager. From the information so gained it became possible to group the bulk of the selling force into three large categories—the good, the poor and the average salesclerks.

	Sec. Man.	
To M	Buyer	Dept.

Please send to ..... of the Employment Office, on this form, not later than....., 19... three lists of

I. List below the names of as many salesclerks as you know who are now employees of R. H. Macy & Co., Inc., and who represent, in your opinion, the most capable and desirable type of salesclerk.

[illegible][illegible]

III. Place below a list of salesclerks who are, in your opinion, neither very good nor very poor; the ordinary type of salesclerk who is satisfactory but not exceptional.

No.	NAME	Remarks

This loose form of rating sheet was used in preference to a more rigid and detailed form for the following reasons:

(1) The department heads of the organization under consideration are to a great degree free agents. It was essential for the return of the largest number of these sheets with a minimum amount of energy and friction entailed, that the form be filled out with the utmost ease.

(2) The department heads are not an academic group—the median intelligence lower than that of such a group—they had consistently shown an aversion to any sort of analytic thinking or patient arrangement of groups into rank order, and such thinking, if forced upon them would be of a doubtful reliability. It seemed on the whole to be wisest to use a spontaneous method of rating—the more so that it would have been impractical to test the entire force of salesclerks. Therefore, by this method of rating, the testing was confined to those individuals who were spontaneously recalled in the minds of the several department heads as meriting good, bad or indifferent ratings.

The results of this method bore out the choice of the course. The forms were returned on time and with a minimum of follow up work. The large body of salesclerks was rated and the undertaking did not lose in prestige by being called super-scientific, impractical, etc.

Department, . . . . .

[illegible]

Fourth, those rated on the whole as average but with one or more judgments that they were *above* average.

*Fifth*, miscellaneous groups of individuals about whom there was no common and consistent type of judgment—those rated as both good and poor, for instance. These were wholly omitted from the experiment because of the inconsistency with which they were rated.

*Clerical.* The other large group of employees engaged in work of generally similar character were the clerical workers.

For the same reasons and in the same fashion as applied to salesclerks ratings of the clerical workers were obtained previous to testing for the three large clerical departments—the Audit, the Mail Order and the Complaint Department. Unlike the salesclerks, however, there was no common production record by which the judgments of department heads could be controlled, the clerks being engaged in a multiplicity of operations, from complicated statistical work and bookkeeping, to counting of checks, filing and comptometer operating. Moreover in most cases the clerks were well known to only one person, the local supervisor, only superficially known to the actual department head and very occasionally known to as many as three persons.

Raters were therefore instructed to rate only those clerks whom they knew and the rates taken at their face value, with the precaution that such clerks about whom there was a wide divergence of opinion were excluded from the possible subjects for experiment, and only such clerks included about whom at least two people were agreed.

The following scale indicates the manner in which the clerks were grouped.

1. Rated as first class by two or more—no dissenting opinion.
2. Rated as first by two—2nd by 1.
3. Rated as first by one—2nd by 1.
4. Rated as second by two—1st by 1.
5. Rated as second by two or three—no dissenting opinion.
6. Rated as third by one—2nd by 2.
7. Rated as three by one and 2nd by 1.
8. Rated as three by two and 2nd by 1.
9. Rated as three by three or two—no dissenting opinion.

In obtaining ratings throughout the course of this work, it was noted that there was a very marked reluctance to say that an employee was third rate or unsatisfactory, for the reason that the executive giving such a rating laid himself open to the question of why he retained such a person. It seemed fair, therefore, to infer that all clerks whose total rating was more than seven, could be included among the

third rate clerks, and they were so considered. This was also done because the number nine group alone made up a very inconsiderable number. The ratings of *good* clerks, if unanimous, were significant; of the poor and mediocre clerks, less so because of the hesitancy to call any one poor, and of a tendency to underrate if the clerk was unknown or doing unimportant work, or if but newly employed.

However, these ratings, such as they were, served to indicate three groups, the very good (1) the poor (7-9) and the generally average clerks who were rated between these two extremes.

*Tests.* Through the courtesy of Prof. E. L. Thorndike and the National Research Council Committee sets of the N. R. C. Tests, the precursors of the National Intelligence Tests were made available for the purposes of this investigation.

These tests were:

- |                                |                          |
|--------------------------------|--------------------------|
| 1. Verbal A which consisted of | 3. Verbal B              |
| 1. Arithmetic                  | 1. Computation           |
| 2. Directions                  | 2. Vocabulary            |
| 3. Sentences                   | 3. Sentence Completion   |
| 4. Synonym Antonym             | 4. Disarranged Sentences |
| 5. Judgment                    | 5. Logical Selection     |
| 2. Non Verbal A                | 4. Non Verbal B          |
| 1. Picture Completion          | 1. Copying Designs       |
| 2. Series Completion           | 2. Pictorial Sequence    |
| 3. Comparison                  | 3. Pictorial Identities  |
| 4. Symbol Digit                | 4. Recognitive Memories  |

In addition the following tests were also used,

<sup>1</sup> Sentence Completion 10' and 5' time limits—called T1A  
 Arithmetic—Woody McCall Mixed Fundamentals—called T2

Rearrangement of Animals—An abbreviated form of the writer's test published in June, 1918, *Journal of Applied Psychology*—called T3

Woodworth Wells Mixed Relations

Woodworth Wells Mixed Relations changed to an underlining test—called T5

Woodworth Wells Opposites

Woodworth Wells Opposites changed to an underlining test—called T4

Woodworth Wells—Number group checking

Woodworth Wells—Cancellation

---

<sup>1</sup> This test is a scale of 25 Trabue Sentences, arranged by the writer, for which norms from approximately 20,000 adults are at hand.



*Testing.* Directions for performing the tests were identical for every individual.

The tests were given to groups in the laboratory sketched above. The testing was always in the early morning—from 9 to 10:30 A. M. Meetings did not last longer than 90 minutes. Subjects were recalled for a second period if the tests were not completed in the first.

Before beginning any testing, the purpose and method of the work was carefully explained to each group as simply as possible. Discussion was perfectly free and testing was never begun without good feeling and rapport between the subjects and the experimenter. Pains were taken to dispel any preliminary nervousness, and to see that subjects took the tests comfortably. Both sexes and all ages from 17 to 60 were represented. Of the salesclerks there were sellers of every type of merchandise from upholstery and women's suits to shoes, veiling, gloves and notions. Every phase of clerical work at which the employees of this establishment were engaged was likewise represented in the test groups. There were skilled typists and stenographers, statistical workers, bookkeepers, billing clerks, comptometer operators, check counters, filers, keepers of simple records, etc., etc.

For the initial phase of the experiment, in which a rough evaluation of the tests as to their sensitivity to different degrees of abilities was sought, only the extremes of both groups were used; that is, of the salesclerks, those were tested who were consistently classed as good and as poor; of the clerical workers, those who, according to the classification scheme indicated above, were classified as 1, and those who fell within the 7 to 9 classifications.

The average workers of each group were made use of in a later phase of the experiment.

#### *Correlation Formula and Statistical Methods*

Several correlation methods were used in the course of this work.

1. The method of Unlike Signed Pairs—discussed in Thorndike's *Mental and Social Measurements*, p. 162, pp. 170-1—and called in tabulations and discussions in this paper the U Formula.

2. Pearson Biserial R.—a formula which determines the correlation when one variable is measured and continuous, the other unmeasured and alternative—and which closely approximates the  $r = \frac{\sum xy}{n\sigma x \sigma y}$  see *Biometrika*, Vol. VII, 1909, p. 96.

3. Finally the orthodox Pearson  $r = \frac{\sum xy}{n\sigma x \sigma y}$  was used in

the last phase of the experiment, in partial correlation and regression in order to weight the significant tests. When this formula was used the value 1 was given to a quality i.e., good salesmanship, and 0 to the absence of it.<sup>2</sup>

The other formulas do not necessitate so arbitrary an assignment of values.

*Correlation Coefficients—Salesclerks.* Table I below represents the coefficients which were obtained from the salesclerks when the good and the poor clerks were tested with a series of tests. The number of clerks tested, the names of the tests and the statistical method used to obtain the coefficients are all indicated.

TABLE I

		GOOD AND POOR SALESCLERKS		r by U Formula
Number of Cases	Tests			
45	N R C - N V A	Total		— .09
46	" " " " " 1	Picture Completion		— .39
44	" " " " " 2	Series Completion		+ .22
44	" " " " " 3	Comparison 3'		— .59
44	" " " " " 3	Comparison 2'		.83
43	" " " " " 4	Symbol Digit		+ .25
44	" " " " " 5	Form Combination		— .37
44	" " " V A	Total		— .59
43	" " " " " 1	Math.		— .16
44	" " " " " 2	Directions		— .66
43	" " " " " 3	Sentences		— .79
44	" " " " " 4	Synonym Antonym		— .22
45	" " " " " 5	Common Sense		— .79
43	" " " " " 6	Analogies		— .54
46	" " " N V B	Total		+ .16
48	" " " " " 1	Copying Designs		— .06
46	" " " " " 2	Pictorial Sequence		+ .16
48	" " " " " 3	Pictorial Identities		.00
45	" " " " " 4	Recognitive Memories		+ .51
48	" " " V B	Total		— .06
49	" " " " " 1	Computation		— .16
49	" " " " " 2	Vocabulary		.00
49	" " " " " 3	Sentence Completion		— .34
49	" " " " " 4	Disarranged Sentences		— .28
49	" " " " " 5	Logical Selection		.00
48	T 1 A	Sentence Completion 10'		— .13
48	T 1 A	Sentence Completion 5'		— .37
33	T 2	Arith.		— .16
49	W. W.	Opposites		— .16
47	T 4 W.	W. Opp. Adapted		— .43
48	W. W.	Mixed Relations		— .37
48	T 5 W.	W. M. R. Adapted		— .46
48	T 3	Rear. Animals		— .56

<sup>2</sup> This method of statistical treatment was suggested by Prof. T. L. Kelly, now at Leland Stanford University.

Some of these coefficients are markedly high, also they are, almost without exception, negative.

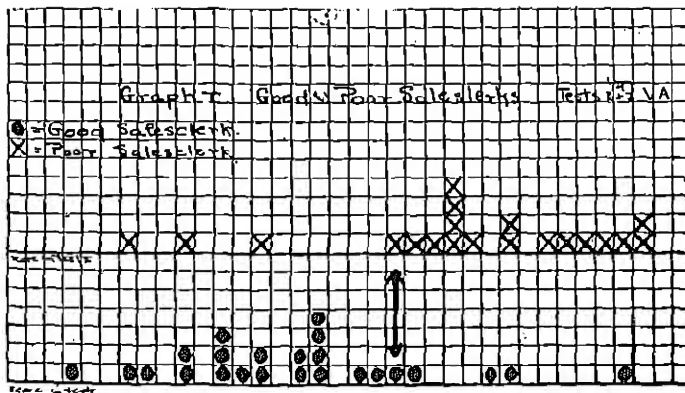
Table II presents coefficients which were obtained from combining in various groups the tests which singly gave the highest coefficients. (The coefficient for NVA3—2' time limit had not been computed at the time these groupings were made and the coefficients of Table II calculated.)

The general tendency, it will be noticed, is to increase the correlation coefficient by such combinations of tests.

TABLE II

## GOOD AND POOR SALESCLERKS' TESTS COMBINED

Number of Cases	Test	r by U Formula
41	Sum V A 3 N V A 3 — 3'	— .66
39	" " " 2-3-5-6 N V A 3 — 3'	— .69
43	" " " 3-5	— .79
43	" " " 2 and 3	— .84
44	" " " 2 and 5	— .78
42	" " " 2 and 6	— .56
44	" " " 2 and N V A 3 — 3'	— .66
42	" " " 3 and 6	— .64
43	" " " 2-3-5	— .84
42	" " " 5 and 6	— .51
44	" " " 5 and N V A 3 — 3'	— .61
39	" " " 6 and N V A 3 — 3'	— .75
42	" " " 2-3-6	— .79
40	" " " 2-3 N V A 3 — 3'	— .66
42	" " " 2-5-6	— .63
40	" " " 2-5 N V A 3 — 3'	— .66



Graphically represented these coefficients take on a more concrete meaning.

Let us consider for a moment the practical implication of such a situation as Graph I represents. Suppose that the forty odd people represented by the sum of the crosses and the rows of dots were, without any distinguishing mark, seated in a room. By means of tests which are represented in Graph I (using the position of the arrow as a reference point) an individual would be able to determine which of the group had been successful and which unsuccessful as salesclerks, and would be in error only 8 times out of the forty-four cases.

*Clerical.* The clerical workers, good and poor, represented by classifications 1 and 7-9 were tested with the following tests,

V A—All  
 N V A—All  
 W. W.—Cancellation  
 W. W.—No. gr. checking  
 Rearrangement of Animals—T3

Table III which follows presents the coefficients which these tests gave, the formula in use being again that of unlike signed pairs.

TABLE III  
 GOOD AND POOR CLERICALS

Number of Cases	Test	r by U Formula
59	N R C - V A	.34
61	" " " " " 1 Arithmetic	.09
61	" " " " " 2 Directions	.00
60	" " " " " 3 Sentences	.31
58	" " " " " 4 Synonym Antonym	.28
60	" " " " " 5 Judgment	.37
50	" " " " " 6 Analogies	.28
48	" " " N V A	.31
52	" " " " " 1 Picture Completion	.13
52	" " " " " 2 Series Completion	.22
51	" " " " " 3 Comparison 2'	.45
50	" " " " " 4 Symbol Digit	.37
52	" " " " " 5 Form Combination	.00
34	No. Gr. Ch.	.31
35	Cancel. .	— .13
57	T 3	.19

It is evident at once that the ratios are not as high as those obtained for the salesclerks, the highest being .45 for N V A 3—the Comparison Test.<sup>3</sup> However, it is also immediately evident that these ratios are all positive, with the one exception of Cancellation. Now it will be remembered that the N V A and V A tests were the tests which gave the highest correlations for salesclerks but that these ratios were almost all consistently *negative*.

It seems fair to assume that had it been possible to secure clerical workers rated with the freedom from error of the salesclerks' classifications, and in view of the fact that a positive relationship is indicated by the plus ratios, that positive correlations perceptibly higher than those actually found would have been obtained. However, this is merely an assumption; what is apparent is the tendency for salesclerks and clericals to pull in distinctly opposite directions, so that we have a difference between trades as distinct as the differences within a trade, if not more so.

This is a not unimportant consideration, for the primary interest of an Employment Office is not in deciding with what degrees of excellency an unknown individual will perform a given task, but to decide at what task to place a given individual.

Concretely, Mary Jones, age 18, without experience or training wants a job. The Employment Manager wants to know whether to make Mary Jones a salesclerk or a clerical. To him whether Mary Jones will rank fifth or first as a salesclerk is a matter of academic interest only. How good she will be is a matter she must show by performance, since in industry as elsewhere reward comes not in anticipation of work but follows it. Wages in a large industry are fairly standardized at employment time, and are modified not for potential ability but only for previous experience. They become really differentiated for individuals only with the history of that individual's work.

In order to express numerically the contrary tendencies of salesclerks and clerical workers, the two groups were thrown together and an attempt was made to separate from this heterogeneous group those who were called good salesclerks and good clericals.

Such a combination seemed to have two virtues, (1) the reliability of the coefficients obtained would be increased because the size of the group would be doubled; (2) the hetero-

---

<sup>3</sup> It will be remembered that this test gave the highest correlation coefficient, but a negative one, when used with salesclerks.

genity of the enlarged group more nearly approximated actual working conditions. Candidates for employment as they present themselves at the office which these tests were to serve are not divided off into sales or clerical workers and often have no strong preference for any special kind of work, but must be classified and selected for various jobs by the individuals who interview them.

This combination of both types of workers was made in two ways, first all sales and all clericals were combined, and selection made, for (1) good salesclerks, (2) good clerical workers. The second grouping was as follows. When good salesclerks were being selected the unsatisfactory clerical workers were omitted from the total group of sales and clericals for the reason that there did not seem to be any ground to consider them as undesirable candidates for selling.

All that was known was that they were undesirable for clerical work. Of course it may be argued that the retention of the good clerical workers was likewise ungrounded; that there was no reason why they should not make good salesclerks, even if they were good clerical workers. This is doubtless true. At the same time, considering again the practical aspect, they were actually non-sellers, the negative-positive correlation findings seemed to indicate that also potentially they might be non-sellers, and it seemed to be of interest to see what would happen to the coefficients already obtained when such a group as this under discussion was made.

When the selection for good clericals was made, the unsatisfactory salesclerks were omitted for the same reasons given for omitting unsatisfactory clerical workers when selecting for salesclerks.

Tables IV and V present the coefficients which were obtained from this double grouping. The coefficients which were obtained from the original pure sales and pure clerical groups are also given for comparison, so it may be seen how the coefficients are affected by the increased number of subjects and the different kinds of groups.

Apparently, the elimination of the bad clericals, in the case of the selection for salesclerks, was justified, for when good salesclerks were selected from the heterogeneous group of all salesclerks and all clericals correlations were lower than those obtained with the salesclerks only. However when the poor clericals are omitted from this heterogeneous group and selection then made for good salesclerks the situation appears to have cleared again, for the coefficients regain their original character occasionally being slightly higher or slightly lower

TABLE IV  
SELECTING GOOD SALESLERKS

Test	Number	All Sales and All Clericals		All Sales and Good Clericals only		Good and Bad Sales Only		Number
		r by U Formula	r by U Formula	r by U Formula	r by U Formula	r by U Formula	r by U Formula	
N R C - V A	102	-.31	-.56			-.59		44
" " " " 1	104	-.22	-.28			-.16		43
" " " " 2	105	-.43	-.56			-.66		44
" " " " 3	103	-.59	-.75			-.79		43
" " " " 4	102	-.12	-.22			-.22		44
" " " " 5	105	-.48	-.64			-.79		45
" " " " 6	103	-.56	-.69			-.54		43
" " " N V A 3-2'	95	-.96	-.93			-.83		44
Rear An.-T3	106	-.19	-.31			-.56		48
V A 3-5 N V A 3-2'	93	-.82	-.91			.....		..
V A 3-5-6 N V A 3-2'	92	-.92	-.94			.....		..
V A 2-3-5-6 N V A 3-2'	92	-.69	-.90			.....		..

TABLE V  
SELECTING GOOD CLERICALS

Test	All Clericals And All Sales		All Clericals and Good Sales Only		Good and Bad Clericals only	
	Number	r by U Formula	Number	r by U Formula	Number	r by U Formula
NRC-V A	102	+.28		+.45	58	+.34
" " " 1	104	+.10		+.16	61	+.09
" " " 2	105	+.03		+.19	61	+.00
" " " 3	103	+.31		+.54	60	+.31
" " " 4	102	+.18		+.27	58	+.28
" " " 5	105	+.54		+.61	60	+.37
" " " 6	103	+.31		+.42	60	+.28
NRC-NV A 3-2'	95	+.66		+.64	51	+.45
Rear An-T3	106	+.16		+.34	58	+.19
V A 3-5 NV A 3-2'	93	+.54		+.71	54	.....
V A 3-5-6 NV A 3-2'	92	+.56		+.79	..	.....
V A 2-3-5-6 NV A 3-2'	92	+.54		+.77	..	.....



than in the original classification. Their significance is however increased, since their P. E.'s are decreased, because of the increased size of the group from which they were obtained.

When good clericals were selected the combining of all salesclerks and all clericals does not seem to affect the coefficients markedly, except to raise them in several instances. With the clericals, it will be remembered, the waters were originally muddy, so that an addition of several more misplacements would probably not have the same disturbing effect as with the salesclerks, where the original groupings were markedly clear.

However, when from the total group the unsatisfactory salesclerks are withdrawn and selection then made for good clerical workers, the coefficients are, with the exception of one case where the ratio remains practically the same, raised to a marked degree, and again we seem to have a justification for having excluded the poor salesclerks from this group.

Not only are the coefficients raised but in some cases they become large enough to promise a fairly satisfactory basis for the selection of clerical workers, something which the size of the coefficients obtained with the original group of clericals only did not warrant.

*Final Phase of the Experiment.* Having, in the manner above described, found several tests which seemed sensitive to selling and clerical ability, it became necessary, in order to employ these tests for the guidance of the Employment Office, to evaluate each test, in accordance with its importance in the group of tests. In other words, it was necessary to determine the regression equation for the whole group of tests, and the best weighting for each test in that group.

Most of the tests however which gave significant coefficients were in the group of the National Research Council Tests. These tests had been constructed for use in schools and their content was not especially adapted for industrial use.

People who are looking for a job usually object to being asked who Black Beauty or Thomas Jefferson are and are apt to ask in response "Is this a school we are in?" Even if this response is not audible, such questions do not result in a very favorable state of mind, and since the success of industrial tests is as dependent upon their reception by the people who take the tests as upon the prophetic accuracy of the tests themselves, new tests were constructed, in form

the same as certain of the N R C Tests, in substance different. The atmosphere savored more of calico, pins and sealing wax and less of George Washington, Abraham Lincoln and the states of the Atlantic Seaboard.

The test questions were all placed on one side of a sheet of  $8\frac{1}{2} \times 11$  paper, on the reverse side of which was a dotted line for the name, and several sample questions of the test itself, the samples being of course so simple as to be self-explanatory. The questions on the test paper were arranged roughly in an order of increasing difficulty, the earliest questions being scarcely more difficult than the sample questions.

A term used by Link, "shock absorber," expresses a perfectly sound principle. Non-academic groups have a very strong aversion not only to academic problems, but mental gymnastics of any kind, and it is much better for the prestige of the test to get such people started on primer tasks and then imperceptibly involved in the more difficult tasks, than immediately to frighten and make them hostile with the more difficult questions.

It was necessary, of course, after the new tests had been made to ascertain if they would act in the same way as the tests they were meant to replace, that is, give the same characteristic correlation coefficients.

This called for more testing. It will be recalled that in originally getting classifications for the salesclerks and clerical workers there were several groups, the very good, the very poor, the average, and the slightly better than average. Of these groups the extremes, good and poor of both workers had been tested in the original phase of the experiment. The clericals and salesclerks who were rated as average had not been tested.

It seemed unwise to test the original groups with the new tests, because, having spent some three hours in test taking they had had an amount of practice which made them no longer naïve, and also because it was inadvisable to take those who had already lost a considerable period of working time away from their departments for a further period.

Consequently the new tests were standardized on the group of untested workers, those who made records of average clerks, or in some cases somewhat better, but who were neither absolutely first rate, nor third rate workers. One hundred salesclerks were tested and 43 clerical workers.

The new tests were modifications of the N R C Common Sense—called in the Tables T 9, N R C Directions—T 8, N R C Comparison—T 13, N R C Sentences—T 12. In addition Completion Test T 1 A—10' time limit—was used, the adapted form of Woodworth Wells Mixed Relations—T 5, and the Animal Rearrangement Test—T 3.

The formula which was first used to evaluate these tests was the Pearson biserial  $r$ , the clerical workers,<sup>4</sup> 43 in number being considered the special group and the 100 salesclerks being considered non-clerical workers.

The coefficients obtained with these tests and this group of workers is given below in Table VI.

TABLE VI

	Test	Pearson Biserial $r$	Number
T 1 A	Completion.....	.19	143
T 8	Directions.....	.63	143
T 9	Sentences.....	.45	143
T 12	Judgment.....	.55	143
T 13	Comparison.....	.74	143
T 3	Rearrangement of Animals.....	.11	143
T 5	Mixed Relations.....	.51	143

With the exception of T 3 and T 1 A the coefficients are high and the performance of the two groups, sales and clerical, is consistent with the performance of the groups previously tested, the salesclerks giving low scores, the clericals high scores.

T 3 and T 1 A could therefore be omitted from any final group of tests. It was decided, in spite of the size of the coefficient of Test 5, which is W. W. Mixed Relations, to leave it out of consideration also, because of the difficulty always encountered in explaining the task and having it understood by those tested. With the other tests it was simply necessary for the subjects to see and do the sample problems, without other explanations from the experimenter. For the sake of economy in time and effort in the administration of the tests, when they should be used as part of the employment process, it seemed wise to discard this test, which had to be explained to the majority of subjects.

<sup>4</sup> It was at the time impossible to test more clerical workers because of the Christmas season rush.

The next step was to find the regression equation for the group of tests which were retained and the best weighting for each test.

The formulae in partial correlation and tests weightings are derived from the Pearson coefficient  $\frac{\Sigma xy}{n\sigma x \sigma y}$ . The coefficients in Table VI are those of the Pearson biserial  $r$ , which closely approximates this  $r$ , but which is not identical with it.

As stated previously, the  $\frac{\Sigma xy}{n\sigma x \sigma y}$  values for the new tests were found by giving the presence of sales ability the value 1, the absence of it or as was actually the case, the presence of clerical ability the value 0, the test scores forming the values of the second variable.

The regression equation was thus determined for 4 variables and the criterion, the variables being the scores in the Directions, Comparison, Judgment and Sentence Tests. The weighting of the Sentence Test however was so very inconsiderable that it was deemed inadvisable to retain it in the group and consequently a second regression equation was calculated in which there were only 3 variables and the criterion.

The following regression equation was obtained using the variables:

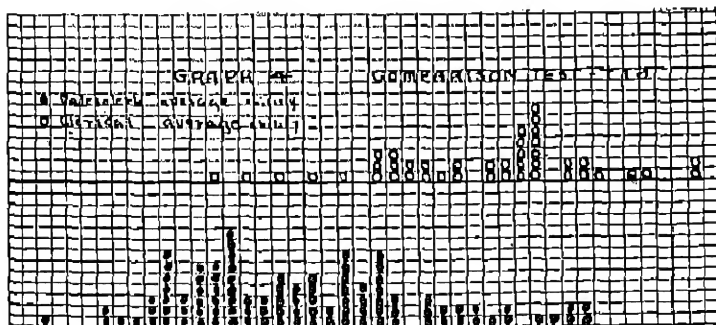
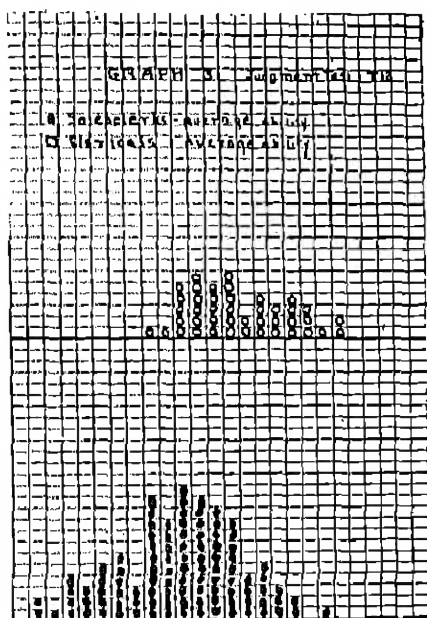
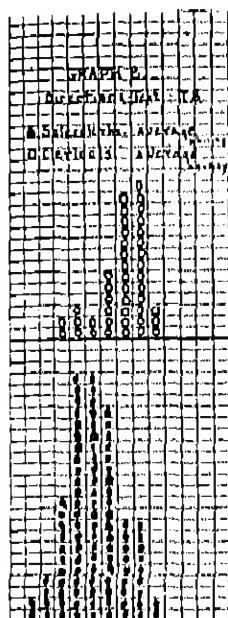
$$\begin{array}{ll} X_1 & \text{The criterion —} \\ X_2 & \text{Comparison Test —} \\ X_3 & \text{Direction Test —} \\ X_4 & \text{Judgment Test —} \\ X_2 & = .025 X_1 + .065 X_3 - .007 X_4 - .85^6 \\ R & \text{for this equation} = .63 \\ \text{Biserial } r & = .74 \end{array}$$

The manner in which, by weighting the tests in accordance with the regression equation above stated the division between the two groups, sales and clericals is emphasized is indicated by the following graphs.

Graphs 2, 3, 4 represent the performance of the sales and clerical groups in single tests—T 8, T 12, T 13. The tendency for clericals to do better work is evident to some degree. A great number of misplacements would occur however if estimates of ability were made by these graphs as they stand.

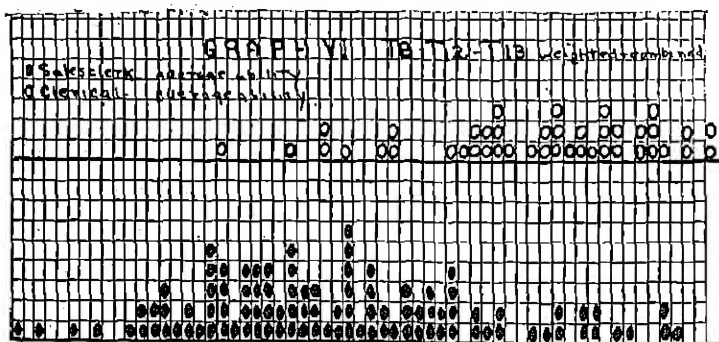
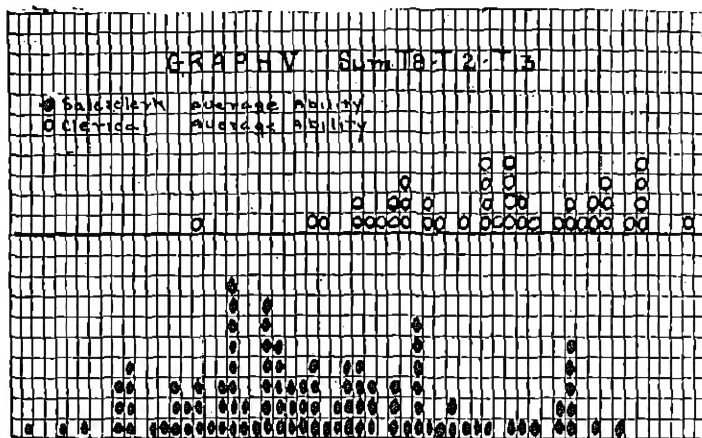
---

<sup>6</sup> This equation is expressed in gross scores.



Graph V is the sum of these tests. Here the two groups show a marked tendency to separate, but there is considerable overlapping.

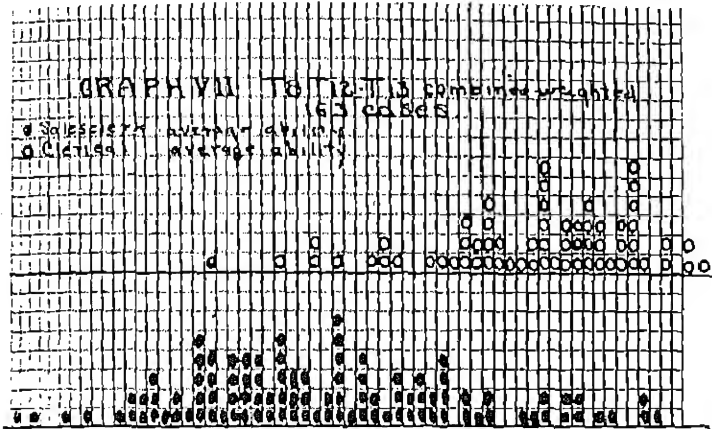
Graph VI represents the tests weighed and combined in accordance with the regression equation.



The overlapping has been diminished—and the salesclerks form a fairly clear characteristic group, the clericals as distinctive a group.

After these regressions and other calculations had been made for the 43 clerical workers, it became possible to test 20 more clerical workers. It seemed inadvisable to recalculate on the basis of the enlarged group the weightings as originally found, but it did seem of interest to see what effect adding these 20 cases would have on the coefficient, when the tests were weighted in accordance with the values that had been obtained from the smaller group.

Pearson biserial  $r$  was raised from .74 with 143 cases to .83 with 163 cases. Graph VII shows the increased emphasis on the separation of the two groups.



*Applications.* All applicants for employment, during the time in which the experimental work above described was being carried on, were being tested with the Completion Scale which has been previously mentioned, and which served to indicate mental defectives and subnormal individuals. Upon the findings discussed above the group of tests T 8, T 12 and T 13 were added to the test already in use and all applicants who were considered for either selling or clerical workers were tested with this group, and placement made as far as possible in accordance with the characteristic range into which the test scores fell, whether sales or clerical. The Completion Scale was still retained and in every case given to the applicant before the group of tests. Thus the fact that a low score, or one in the range characteristic of salesclerks was due to general mental inferiority, and not other causes was guarded against.

The tests have been in use for some time. In about 50% of the cases interviewers have applicants tested before placement, the other 50% of the cases are placed and then tested—the placement being subject to revision when the test findings strongly contradict the placement.

Applicants are tested generally by a clerical worker, who works under the direction of the writer and has been specially

trained to give and score the tests. As many as 160 applicants have been tested in one day. Applicants are tested singly, or in groups up to 10.

The tests are also used in cases of personnel adjustment, transfers, selection of candidates for training classes, etc.

A careful follow up by rating and production record is being carried on in order to check the test records with actual work. These are not yet in such a form that findings can be presented statistically, but on the whole, from a rough survey of the evidence at hand and from isolated cases which have come in for special attention the findings reported above tend to be borne out.

*Summary.* In a large department store methods of obtaining reliable criteria for work ability of salesclerks and clerical workers were carefully developed: Workers were grouped into three classifications, very good, poor and average, and tested with a series of mental tests; only those workers being tested who were consistently placed in one of these three classifications by all available criteria.

The good and poor salesclerks when tested gave negative coefficients of correlation, some markedly high. Similar groups of clerical workers likewise tested gave positive correlations, not so high.

When the two groups of workers, sales and clerical, were combined, the good workers of each type could be selected from the heterogeneous group by test marks, generally with accuracy as great or greater than from the homogeneous group. This was especially true if the poor workers of one group were omitted when selecting for the good of the other.

New tests were devised which were similar in form to those that gave the highest correlations. These tests differed in content by being less academic, more industrial. These and several tests in their original form were given to a group of 100 average salesclerks and 43 average clerical workers. The relationship between the two groups which had obtained with the original tests and the original groups tested held with the new tests and the new groups.

The regression equation was determined for the three tests which, taken singly, gave the best correlations—T 8, a Direction Test, T 12, a Judgment Test and T 13, a Comparison Test. The best weightings for these tests when combined were thus determined.

Biserial  $r$  for this group of tests and the group of 143 workers was .74. Twenty additional cases, clerical workers, raised the coefficient to .83.



These tests are now being used in connection with the employment of sales and clerical workers.

At a later date it is hoped to present data resulting from the follow up of the work of employees so tested.

In presenting this record of the work toward the development of tests for special ability, the writer is not unaware of the fact that much of the procedure and handling of the data varies from orthodox laboratory methods. Ideal laboratory conditions, are however rarely, if ever, possible to an industrial psychologist, who must make what shift he can. The exigencies of the occasion are therefore offered as an excuse for the departures from the orthodox. That in spite of necessarily improvised methods certain relationships were seen to hold consistently true, might in itself be offered as an "end justifying the means" excuse.

Also the subject of the negative correlations which were found, their why and wherefore, is a matter of some interest and importance which the present paper does not touch.

The findings themselves are presented, for comment and criticism, and with the thought that, in the light of the present scarcity of work on tests for special abilities, the greater scarcity of literature on the same subject, and a certain feeling of discouragement which has recently become associated with work in this field, the history of this work would possess some interest.

---

NOTE.—It should be noted in connection with the individuals tested in this experiment, that they possessed one attribute in addition to that of ability or disability at a special job, that attribute being *permanency* as workers. The workers were rated in June, sales records were begun from May, the testing lasted into January of the following year and ratings were unobtainable for workers who had not been employed for some length of time previous to June. Consequently only such individuals could have been included among those tested as were fairly permanent workers. Stability is as important a quality in a satisfactory employee as ability itself.

It is conceivable that any superior individual, making in the tests a score beyond the range of both clerical workers or salesclerks would do excellent work at either job. It is however (since promotion can of necessity come only to a limited number), highly inconceivable that such an individual would remain for more than the briefest period at such work as department store salesclerks or clerical workers are called upon to perform, day after day, and such individuals could hardly be counted on to make up the bulk of the 5,000 workers who carry on the business of such an organization.

The median intelligence of the workers in this store approximates, (from data not yet published), the degree of intelligence which was found characteristic in the Army Tests, and it is for such individuals that the standards here set forth hold true.

## ESTIMATING INTELLIGENCE BY MEANS OF PRINTED PHOTOGRAPHS

---

By L. DEWEY ANDERSON, Bureau of Personnel Research, Carnegie  
Institute of Technology

---

### PURPOSE

The purpose of this investigation was (1) to determine the reliability of photographs for indicating the intelligence of strangers, and (2) to seek out differences in the ability of various persons to judge intelligence by this means. The results of the study might be applicable to problems of selection arising in an employment office.

### METHOD

The photographs used were prints (from cuts) in a handbook published by a large department store company. They represented the faces of 69 employees of that company. These persons consisted of superintendents, buyers, managers and assistants.

In the fall of 1920 the same 69 persons were given the Bureau of Personnel Research adaptation of the army intelligence test, and the scores in this test were used as the intelligence criterion.

The photographs were all bound in the company annual, or handbook, but the printed matter on each page was concealed by a paper shield. The person judging the photographs, therefore, had to finger through the book, picking out the individuals, and designating them by the number which was given to each, to indicate their intelligence rank as described below.

Twelve judges passed upon the photographs. These judges were selected from a nearly homogeneous group of graduate students and instructors in psychology and there was very little chance of other influences entering into the judgments of intelligence given.

The instructions were as follows:

### INSTRUCTIONS

"These photographs will introduce you to the 69 managers, buyers and assistants of the ——— Company. You are asked

# ESTIMATING INTELLIGENCE BY MEANS OF PHOTOGRAPHS 153

to estimate the intelligence of these people by a careful study of their photographs.

"First, scan through the book and get an impression of the general character of this group. Then pick out the following:

- (1) The seven most intelligent persons.
- (2) The seven least intelligent persons.
- (3) The fourteen persons who are superior, but not as good as the best seven.
- (4) The fourteen persons who are inferior, but not as poor as the worst seven.

"The reason for the grouping by sevens is found in the fact that there are about seven individuals in each decile. The highest 10 per cent, the next 20 per cent, the lowest 10 per cent and the next lowest 20 per cent, are thus sought. This would leave an undifferentiated middle group of 40 per cent. The entire arrangement is thus intended simply to sift out the extremes."

TABLE I  
SUCCESS OF VARIOUS JUDGES IN ESTIMATING INTELLIGENCE BY  
PHOTOGRAPHS

Judge	No. of Absolutely Right Judgments	No. of Absolutely Wrong Judgments	No. of Judgments on the Correct Side	No. of Judgments on the Incorrect Side
A	12	1	27	15
B	7	1	23	19
C	7	1	25	17
D	4	1	24	19
E	6	2	19	21
F	10	1	26	16
G	8	0	22	20
H	10	2	22	20
I	8	1	20	22
J	7	1	20	22
K	7	1	22	20
L	10	2	22	20
	96	14	271	231

In this table "absolutely right" means that the man's photograph was assigned to the *exact group* into which the intelligence test placed him. "Absolutely wrong" means that the very superior man was marked very inferior or the very inferior man was marked very superior. Placing the inferior man in the very superior group or the superior man in the very inferior group, etc., is indicated in the fifth column. "Judgments on the correct side" are those where superiority to the average has been rightly discriminated and where inferiority to the average has been rightly discriminated.

## ANDERSON

From the above table it will be noted that there are 96 absolutely right judgments and only 14 absolutely wrong judgments. This means that out of 110 judgments, 96 were absolutely right and 14 were absolutely wrong, or placed in the directly opposite group. In the last two columns of the table there are 271 judgments on the correct side and 231 on the wrong side. These data show that the judgments of intelligence based on photographs are, in this experiment, fairly accurate where only the extreme degrees of intelligence as measured by other criteria are involved. If all judgments are considered, i.e., if the middle group that is badly 'judged' is included, there is a very slight tendency to accurate placement.

A product moment correlation was made by assigning values to votes as follows: A judgment as very intelligent was given a rating of 2, a judgment in the least intelligent group a rating of -2, a judgment of superior, a rating of 1 and a judgment of inferior, a rating of -1. The ratings given to each photograph were summed and a product moment correlation of  $+0.27$  with intelligence test scores was secured. The probable error is .07. This correlation shows that there is a positive relationship between photograph ratings and intelligence test scores.

Another correlation was figured by the rank method between the judgments of 31 store managers on each other and the photograph ratings. The managers' judgments were made on the basis of general value to the firm or general ability as a store manager. A correlation coefficient of  $+0.22$  was found. This shows that there is some reliability in the correlation coefficients.

In all of these computations the "average" group is omitted.

## CONCLUSIONS

I. *General Reliability of Photographs*

It seems very doubtful whether for practical purposes intelligence can be gauged by a study of a man's photograph. The correlation of  $+0.27$  between assigned ratings and intelligence as measured by the Bureau of Personnel Research Intelligence Test shows, however, that there is some tendency to discriminate intelligence above mere chance selection. From a study of Table I it was found that 16 men were placed at least once in both the most superior and the most inferior group, 49 people were placed in both the superior and inferior groups, 18 people were placed in both the most intelligent and the

inferior groups and 29 were placed in both the least intelligent and the superior groups. From these results, it appears that selecting employees by a study of photographs would be a very haphazard way of selection. When competent raters place men in both the highest and lowest classes 16 times out of a possible 69 chances, there is little possibility of upholding the practice of photograph rating.

There seemed to be considerable agreement among the raters in regard to some photographs. A point suggested by this agreement is that people in judging a person by his photograph usually pick out a particular feature such as the lustre of the eye or the width of the brow from which to draw their conclusion. In this study it was impossible to check this belief.

## II. Individual Differences

There are considerable individual differences shown in ability to make ratings on intelligence from photographs. One rater made only four absolutely right judgments, while another made twelve. The average number of correct judgments is 8. From Table I, the number of absolutely wrong judgments is 14, thus showing, when compared to a total of 96 absolutely right judgments, that the raters were using some criteria of judging that served to pick out right men. The difference between poor and good raters is not as well marked in judgments that merely rightly place the photographs as in "absolute" judgments. The best record was made by "A," having 27 right judgments and 15 incorrect ones. The poorest records were made by two people, I and J, each having 20 correct and 22 incorrect judgments.

By dividing the number of correct judgments by the total number of judgments a percentage of accuracy was gained for each rater. These percentages follow:

A.....	64%	G.....	52%
B.....	55%	H.....	52%
C.....	59%	I.....	47%
D.....	54%	J.....	47%
E.....	47%	K.....	52%
F.....	62%	L.....	52%

## TWO CASES SHOWING MARKED CHANGE IN I. Q.<sup>1</sup>

By W. T. Root, University of Pittsburgh

Every psychological examiner has, from time to time, certain cases in which testing and retesting even within a brief period of a few months yield decidedly different mental ages and consequently different I. Q.'s. Furthermore many of these cases cannot be traced to clerical errors, to moot questions in scoring, or to the personal equation, which may be a factor of discrepancy when different persons have administered the test at different times.

The cases here presented are typical of a not infrequent performance where school conditions permit individual attention and free grading. Both children came up for examination the second time due to the curiosity of certain teachers who had noted a marked improvement in the character of the work in each case, and also improvement in alertness and grasp. Perhaps the best way to get at these cases will be to give in detail the tests not passed the first time but passed the second time.

### CASE A

Age first examination, 9-0; mental age, 8-4; retardation, 8 months; I. Q., 92. Second examination, age 9-4; mental age, 9-9; acceleration, 5 months; I. Q., 104. Both times year VII was the base year. The first time the following tests were failed: year VIII, 1 (very poor), 2 (nearly performed), 5 (*b* is plus, *a*, *c*, and *d* failed); year IX, 1 (day of week only, given), 2 (nearly correct, 12 and 15 gram weights interchanged), 3 (very slow and inaccurate, not even good approximations, e. g., 25-4 equals 7), 4 and 6 (nearly passed). As IX 5 was the only one of this year passed, the testing would have closed with this year, but for the vocabulary score which carried the examination into the X year. In this year tests 4, 5, and A1 2 were failed. Tests 6 and A1.3 were not given. All of year XII failed.

The second time the examination was given the following tests were passed that were failed the first time: Year VIII, 1 (just scored), 2 (28", 0 errors); year IX, 1, 2 (3 still failed 25-4), 6 (passed easily). The score for X remains unchanged

<sup>1</sup> From School of Childhood, University of Pittsburgh. Stanford Revision of Binet Tests used.

with the exception of A1 2 which is passed (barely) this time, but does not count in the score. In the XII year the interpretation of pictures is now sufficiently good to warrant scoring here.

In this case two factors offer plausible explanation for the advance in the second examination: (1) marked improvement in sight and the clearness of discrimination in reading; (2) rapid academic advance (noted by teachers) and most plausibly attributed to free and semi-individual method of teaching which permits a direct attack on the peculiar weaknesses of a pupil, in this case the three R's especially. A general clearing up in mechanics of reading, and a slight increase in verbal fluency and number combination would account largely for the differences in the two tests.

#### CASE B

Age first examination, 6-8; mental age, 7-6; acceleration 10 months; I. Q., 112. Second examination, age 6-11; mental age, 9-8; acceleration, 33 months; I. Q., 140. In the first examination year V is the basic year; year VI, test 1 is failed; year VII, 6; year VIII, 1 and 6 (vocabulary 16); year IX, all but number 4 are failed, thus closing the examination.

In the second examination tests VI-1 VII-6, and VIII-1 and 6 are passed making VIII the basic year. The vocabulary is 22 words. There is a good performance of the Ball and Field against a complete failure the first time. In year IX tests 5 and 6 are passed with ease while 4 (repeat 4 digits backwards) is again barely passed. In year X, tests 2, 3, 5, and 6 are passed; in year XII, 3 and 7. Tests 2 and 6 (year X) are performed with ease and would probably have been passed the first examination.

In this case the increase in I. Q. can be accounted for plausibly by the factors suggested as causal for case A. Besides these factors which first suggest themselves, several other causes for fluctuation may be outlined tentatively:

1. A school permitting individual advance can make an attack on individual academic weakness, thus attention may be focused upon sentence structure, phonograms and addition combinations if they have proved to be weak. Such drill could conceivably have unintentional but nevertheless indirect effect upon year IX, 3, 5 and 6 (Making change, Using three words in a sentence, and Rhymes, respectively). These three tests correctly performed mean an addition of six years mental age. In the School of Childhood reading is delayed until the second grade, which might readily account for the failure in Rhymes

in the first grade, and passing of the test with ease after drill in *ay*, *ill*, and *ing* endings.

2. Every examiner has noticed that a child may just fail in several allied tests. Chance conditions may make these tests just barely plus on reexamination. Every teacher has also noted that after a rather discouraging plateau, a child may suddenly get a "hunch" in reading or number work. At times understanding seems to act on the *all or none* principle. The emotional attributes of failure and success could also play a part in test performance. (Case A and B, it will be recalled, came up for reexamination because their teachers had noted a marked improvement in their grasp of subject matter.)

3. Superior children are often those who have been brought under adult influence and culture to a marked degree. They possess a peculiar "adultish" ability to answer in a mature way verbal and moral questions, which gives a wide scatter range in testing, making it unsafe to discontinue the test per instructions. For example, all of year IX may be failed while test 2 and 5 year X and test 5 and 7 year XII may be passed. This is precisely what subject B did.

4. Improved physical condition may make an appreciable change in a short time.

Such cases as here described seem to indicate the need of more careful qualitative examination of test responses in terms of individual differences. Also there should be frequent reexamination where the nature of the school freely admits of more than average advance. And finally there should be frequent reexamination of children who have made marked physical improvement from any cause.



## A PRELIMINARY STUDY OF THE CORRELATIONS BETWEEN ESTIMATES OF VOLITIONAL TRAITS AND THE RESULTS FROM THE DOWNEY "WILL- PROFILE"

By G. M. RUCH, Fellow in Education, Stanford University

The results summarized in the tables which follow were gathered in the attempt to compare the estimates of associates with the actual test scores earned in the Downey "Individual Will-Temperament Test" for the twelve traits comprising the scale.

Estimates were secured from two groups of associates, viz., university instructors and students in the same classes, for more than twenty advanced or graduate students. Except in two cases those rated had been known to the judges for at least six months and in most cases the judge and the student rated were fairly intimately acquainted socially as well as in academic associations. The numbers of judges varied from two to five for the faculty group and from four to eleven for the student group. No cases where the numbers of estimates secured were less than two or four, respectively, are included below; thus reducing the numbers concerned to fifteen except where a larger number is definitely indicated. The instructions given to the judges as to the nature of each volitional trait to be rated were taken verbatim for ten of the twelve from Dr. Downey's instructions in her paper, "Some Volitional Patterns Revealed by the Will-Profile" in the *Journal of Experimental Psychology*, Vol. 3, 1920, pages 281-301. The remaining two traits which have been added to a later revision of the scale, viz., "Finality of Judgment" and "Volitional Perseveration" were described by the writer in the instructions. It was suggested that the judges rate the entire group upon trait one before proceeding to the next trait in the hope that greater comparative discrimination would be exercised by those making the estimates.

The correlations were made by the Order-of-Rank method of Spearman corrected to approximate the Pearson Product-Moment formula.

The correlation between the total scores (sum of scores in twelve traits) and scores in Army Alpha for fifteen subjects was found to be  $-0.09$ . For the pooled estimates of faculty judges and Alpha,  $r$  was  $-0.12$ .

TABLE I

SHOWING THE CORRELATIONS BETWEEN THE SCORES FOR THE TWELVE TRAITS AND THE POOLED ESTIMATES OF THE JUDGES FOR THE SAME TRAITS, SUBJECT BY SUBJECT

Subject	Pooled Estimates of Faculty Judges with Test Scores	Pooled Estimates of Students with Test Scores	Pooled Estimates of Faculty with Pooled Estimates of Students
P.....	.26 (3)	.48 (11)	.88
Wy.....	-.06 (3)	.54 (10)	.16
Fe.....	.00 (5)	.31 (10)	-.65
Dar.....	-.15 (3)	-.13 (8)	.87
C.....	.02 (4)	-.13 (10)	.59
Wh.....	.13 (3)	.23 (10)	.53
R.....	-.18 (2)	-.61 (9)	.84
Dav.....	-.12 (2)	.57 (9)	-.04
B.....	.29 (4)	.10 (10)	.76
We.....	-.13 (3)	-.38 (10)	.76
Fa.....	-.11 (4)	.14 (7)	.83
L.....	.17 (4)	-.40 (5)	.76
H.....	-.13 (4)	-.05 (4)	.43
T.....	.03 (2)	.33 (10)	-.30
Deb.....	.10 (3)	.22 (7)	.70
Average..	.01	.08	.47

( ) Number of judges.

TABLE II

CORRELATIONS OF THE ESTIMATES OF FACULTY AND STUDENTS AND THE TEST SCORES AND BETWEEN THE ESTIMATES OF THE TWO SETS OF JUDGES, TRAIT BY TRAIT

Test Number	Trait	Test with Fac. Est.	Test with Stud. Est.	Fac. with Stud. Est.
1.	"Speed of Movement"	.02	.43	.72
2.	"Freedom from Load"	-.09	.27	.28
3.	"Flexibility"	.45	.17	.22 (S)
4.	"Speed of Decision"	-.02	.53	.67
5.	"Motor Impulsion"	.51	.50	.73 (S)
6.	"Assurance"	.23	.37	.86 (S)
7.	"Resistance"	.35	.28	.67 (S)
8.	"Finality of Judgment"	-.11	.07	.45
9.	"Motor Inhibition"	-.33	-.23	.66
10.	"Care for Detail"	.15	-.07	.77 (S)
11.	"Coordination of Impulses"	-.09	.05	.64 (S)
12.	"Volitional Perseveration"	-.26	.05	.72
Average.....		.07	.20	.62

(S) Scoring in part subjective. All others purely objective.

TABLE III

AVERAGE INTER-CORRELATIONS OF THE THREE TYPES OR GROUPS OF TESTS WITHIN THE DOWNEY SCALE, VIZ.: TRAITS 1-4 OR "HAIR-TRIGGER" PATTERN; TRAITS 5-8 OR "WILFUL AGGRESSIVE" PATTERN; AND TRAITS 9-12 OR "SLOW, ACCURATE, TENACIOUS" PATTERN

Pattern	Average Inter-Correlation of the Four Tests Mak- ing up the Pattern
I. "Hair-trigger" (1-4).....	-.005
II. "Wilful Aggressive" (5-8).....	.04
III. "Slow, Accurate, Tenacious" (9-12).....	.06

TABLE IV

THE RELIABILITIES OF ESTIMATES AND TESTS AS OBTAINED BY THE METHOD OF CORRELATING THE SCORES OR ESTIMATES FOR THE ODD-NUMBERED TESTS WITH THOSE FOR THE EVEN-NUMBERED TESTS, THE SCORES BEING THE TRAIT-SCORES

	r	N
Downey Scale.....	-0.15	15
	-0.21	22
Faculty Estimates.....	0.86	15
Students' Estimates.....	0.58	15

From table I it seems evident that the estimates of the judges show no relation to the actual test scores but that they do show a fair amount of agreement between the two sets of judges (average, .47). The students seem at first sight to be somewhat the better judges, but the apparent difference may be due to the fact that the number of students furnishing ratings was considerably greater than the number of instructors.

Table II shows a similar result. Considerable agreement between judges but little or none between estimates and tests is shown. It might be stated that students and instructors have a common concept of the significance of the traits (or rather of certain of the traits) but that these concepts do not correspond with the psychological processes concerned in the tests of the Downey scale. It was feared by the writer that his comparative inexperience in giving the tests, and particularly in scoring, might be responsible for the lack of higher agreement between estimates and tests. However, on the whole, those tests where the scoring is purely objective give poorer results than those where the personal equation does enter to some extent.

The treatment of data as given in table III calls for definite reference to an issue involved which may be entirely unfair to the Downey tests. The assumption has been made in

resorting to the method of average inter-correlations within the three groups of four tests each which have been designated as "patterns" that there should be a fairly high community of psychological processes within such patterns and that there would be a corresponding absence of close relation between tests of different patterns. This may not be a true conclusion, although Dr. Downey has stated that the tests are so grouped, tentatively, in order to bring together tests of the same general type.

Similarly in Table IV the same assumption has been made. The correlation of the odd-against the even-numbered tests really breaks the entire scale into half-scales in which each pattern is represented by two tests rather than four tests. This again may be unfair. However, the nature of two or more of the tests of separate traits makes it impossible to obtain measures of reliability by repeating the tests on another day. The increase in the number of subjects from fifteen to twenty-two lowers the coefficient somewhat rather than raising it.

The extremely small number of subjects used makes it imperative that all criticism of the Downey Will-Profiles be foregone even if the writer were so disposed. These results may possibly have some value and interest to others and it is hoped that they will serve to aid the efforts in a pioneer and inestimably valuable field of research. It is hoped to extend these preliminary findings as soon as possible to a point where greater reliability can be assured for the conclusions which may be drawn.

April 1, 1921.

## PRELIMINARY REPORT ON A GIFTED JUVENILE AUTHOR\*

By LEWIS M. TERMAN and JESSIE C. FENTON

Data on gifted children have been accumulating at Stanford University since 1911. In 1915 one of us published a brief summary of facts relating to 54 children testing above 120 IQ, and certain additional material relating to 31 who had an IQ of 125 or above. (Terman; "The Mental Hygiene of Exceptional Children," *Pedagogical Seminary*, 1915, 22, 529-537.) In 1916, with the assistance of Margaret Hopwood Hubbard, more systematic work was undertaken in the locating, testing, rating of gifted children. In *The Intelligence of School Children* (Houghton Mifflin Co., 1920) fairly extensive data were reported on 59 cases, most of whom tested above 140 IQ. From 1917 to 1919 little was accomplished, owing to the interruption caused by the war, but during the last two years the work has occupied somewhat more than half the time of a research assistant. Our files now contain test records and other data on nearly 200 gifted children, most of them above 140 IQ. These are being followed up from year to year and as many as possible are given re-tests. At present the entire plan of work is being revised preparatory to gathering extensive psychological, educational, social and biological data on 500 to 1,000 of the most gifted children in California. An important feature of the plan is to follow up as many of these cases as possible for a period of at least ten or twenty years. One of the most interesting of the children thus far studied is the gifted juvenile author described in the following pages.

Betty Ford was born in San Francisco, January 21, 1912. Her four grandparents were of Swedish, German-French, English, and Scotch descent, respectively. There is nothing especially remarkable in her family tree. The mother is a woman of more than average intelligence and of considerable musical ability. The father is a physician and the author of the "Ford Stitch," favorably mentioned in standard texts on surgery. Betty is an only child.

---

\*Read before the American Psychological Association, December, 1920.

In Betty's baby-book, kept by her mother, we find the following entries:

Weight at birth, 11 pounds, 15 ounces.

Weight at 20 months, 31 pounds.

At 7 months began to walk about in her "walker."

At 19 months said everything clearly and knew the alphabet.

At 20 months put together block puzzle pictures.

At 22 months enumerated birds up to 12 in a picture.

In view of her intellectual precocity it is interesting to note that Betty's earliest memories, if we can accept her own testimony, date back to the time when she was still fed by bottle, at the age of about 12 months.

Betty's parents say that they first learned of her ability to read when they discovered her, at the age of  $4\frac{1}{2}$  years, reading Heidi, a book of about fourth grade difficulty. At five she read such books as Silas Marner and Charles Lamb's stories. By her eighth birthday she had read approximately 700 books, many of them twice. For the last two years she has read on an average at least eight hours a day. She reads rapidly and with complete absorption. Instead of finishing one book before proceeding to another, she prefers to keep three or four going together. At the age of eight her favorite authors included Shakespeare, Burns, Longfellow, Wordsworth, Scott, Poe, Victor Hugo, Thackeray, Dickens, Cooper, Mark Twain, Barrie, Conan Doyle, Kipling, R. L. Stevenson and George Eliot.

At the age of 30 months Betty began making up little stories about rabbits, frogs, fishes and squirrels. At 33 months she composed little jingles in rhyme. Stories and jingles continued to appear from time to time, but before the sixth birthday only a few were recorded. When she was six years old she was given a typewriter, and with its assistance and stimulus she was soon well launched on her literary career. By the age of eight she had written more than 100 poems and more than 75 stories, together enough to fill a volume of 200 or 250 pages.

We first saw Betty in November, 1919, when she was 7 years,  $10\frac{1}{2}$  months old. Since that time we have spent approximately thirty hours with her, giving her tests, showing her about the University or visiting with her in her home. Six weeks before her eighth birthday she tested 14-10 by the Stanford Revision; IQ 188. Her vocabulary at that time was approximately 13,000 words, which is not far from the median for Stanford University freshmen. Her score on the Army

Beta test was 71, which is about the norm for 14 years. Thus her IQ measured by a non-verbal scale is 175. By the National Intelligence Tests she graded above 15 years (score, Form A 152.8, Form B 123), and by the Terman Group Test considerably higher (score, 151, Median for Grade 12). On the Trabue Completion tests, B and C, her score (17.5) equaled the median for third or fourth year high school pupils. She took the memory span test for digits (oral presentation) with a class of 21 graduate students, beating ten and tying four. Her extreme memory span is 9 digits, direct order, and 7 digits, reversed order. On the Kelley construction test, which was given her by Dr. Kelley, her performance resembled that of a normal child of nine or ten years. On the various types of tests found in the Stanford Revision, the National, the Terman Group and Beta, her performance runs fairly even, though not perfectly so. She is at her best in vocabulary, information, language completion, and memory, and at her worst in tests involving arithmetical processes.

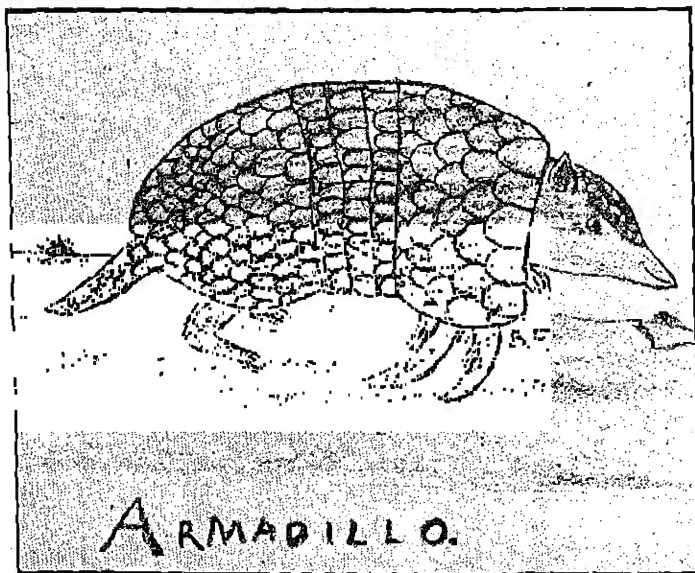
Betty has never attended school, but for nearly two years she has had about two hours a week of private instruction in dancing, French and German. During the last year her mother has given her, intermittently, a small amount of instruction in arithmetic, usually not more than an hour a week. On the Woody tests for addition, subtraction, multiplication and division her scores correspond to the medians for grade 5. On the Kansas Silent Reading test she scores at median for the 11th grade, on the Thorndike handwriting scale at median for 8th grade, on the Starch Grammatical Scale (A) at median for 8th grade, on the Starch Punctuation Scale (A) at median for 6th grade, on the Teachers College Spelling Scale (List 16) at median for 8th grade, on the Grier tests of information in zoology and physiology considerably better than the standards given by Grier for second term high school pupils, and on the Abbott-Trabue test of poetic appreciation at the median for second year college students.

As a test of Betty's poetic ability, five of her poems, selected by a dozen judges as her best, were chosen for comparison with twelve other poems, including the following: three from a published volume of poems written by Stanford students, three by Longfellow (one of them written when the author was 9 years old and one when he was 13), one by Tennyson, one by Blake, one by Blanden, one by Wordsworth, one by Herrick, and a juvenile poem by Shelley. These twelve poems and five of Betty's with names of authors omitted, were arranged for order of merit by 46 Stanford students (chiefly

seniors) enrolled in advance classes in versification, Browning, and composition. In the combined rank order, based on the ratings of the 46 judges, Betty's five poems held ranks 8, 9, 10, 11 and 14. For each of her poems the mean deviation of the 46 rankings from the average rank was between 2.5 and 4.1.

The rank orders based on the combined ratings were as follows:

1. Monterey: Stanford poem by Glenn Arthur Hughes
2. Song: Stanford poem by Dorothy Gunnell Jenkins
3. Shaheen: Stanford poem by James Leo Duff
4. Snowflakes: Longfellow
5. The Eagle: Tennyson
6. The Battle of Lovell Pond: Longfellow (Written at age 13)
7. The Song the Grass Sings: Blanden
8. My Prayer: Betty Ford
9. After the Rain: Betty Ford
10. The Clouds: Betty Ford
11. Joy: Betty Ford
12. March: Wordsworth
13. The Wildflower's Song: Blake
14. The Rainbow: Betty Ford





15. To Violets: Herrick

16. Mr. Finney's Turnip: Longfellow (Written at age 9)

17. The Cat in Distress: Shelley (Written at age 8)

The selection of poems for rating demands a word of explanation. In general, poems were selected which were not too unlike Betty's in content, which were presumably not well known by the judges, and which were relatively brief. Three were selected because they were juveniles. The five of Betty's were selected by about a dozen graduate students as being among her best. Three of the five were judged by Betty as among her best.

It will be noted that the first three places are won by Stanford poems. This may be entirely due to the fact that they were probably known to a good many of the judges, though they are not without considerable merit. Two of the juvenile poems by well known authors are rated at the bottom of the list, but one of the juveniles by Longfellow is given fourth place. We reproduce some examples of the seventeen poems, those of Betty in their original spelling and punctuation.

#### SONG

(Stanford poem. Rank 2)

There is life in the green of the hills,  
There is joy in the mist in the hollow,  
Where the song of the meadow lark thrills,—  
Up, my heart, follow.

There is joy in the song of the sea,  
There is life in the dip of the swallow,  
And the voice of the world calls to me,—  
Up, my heart, follow.

#### THE EAGLE

(Tennyson. Rank 5)

He clasps the crag with hooked hands;  
Close to the sun in lonely lands,  
Ringed with the azure world, he stands.

The wrinkled sea beneath him crawls;  
He watches from his mountain walls,  
And like a thunderbolt he falls.

#### MY PRAYER

(Betty Ford. Rank 8. Written at age 7 years. 6 months)

Oh, Master of fire, Oh Lord of Air,  
Oh God of waters, Hear my prayer,  
Oh, Lord of ground and of stirring trees  
O God of Man and of pleasant breeze  
Dear Father, Let me happy be  
As happy as a growing Tree.

## AFTER THE RAIN

(Betty Ford. Rank 9. Written at age 8 years, 10 months)

Birds carol glad and free,  
 All about joy you see,  
 After the rain!

Each little throat swells glad,  
 None can be ever sad,  
 After the rain!

Moss is a velvet green,  
 Grasses of silken sheen,  
 After the rain!

Blue is the sky above,  
 Everything smiles with love,  
 After the rain!

Birds serenade the sun,  
 Each little songster one,  
 After the rain!

Flowers smile back at him,  
 No one can e'er be grim,  
 After the rain!

And so let Nature prove,  
 In all is God's great love,  
 After the rain!

## THE CLOUDS

(Betty Ford, Rank 10. Written at age 7 years, 3 months)

The Clouds are soft white feather beds  
 On which the fairys lie  
 the fairys all look down on us  
 as they come floating by

but when the clouds grow angry  
 they throw the Fairys down  
 The fairys are so happy  
 to visit at our town.

Ques.

What are the fairys?

Ans.

Raindrops!

THE RAINBOW

(Betty Ford. Rank 14. Written at age 7 years, 3 months)

The rainbow of Diana  
Is shineing in the sky  
The sunbow of Apollo  
has not just yet come by.  
Oh shine again  
thou lovely rainbow  
the dark'ning  
Twilights coming now  
thy brother's waiting  
O'er the sea  
He hath not yet come  
home to thee.

A CAT IN DISTRESS

(Shelley. Rank 17. Written at age 8)

A cat in distress,  
Nothing more nor less:  
Good folks, I must faithfully tell ye,  
As I am a sinner,  
It waits for some dinner  
To stuff out its own little belly.

You would not easily guess  
All the modes of distress  
Which torture the tenants of earth;  
And the various evils,  
Which like so many devils  
Attend the poor souls from their birth.

Some a living require,  
And others desire  
An old fellow out of the way;  
And which is the best  
I leave to be guessed,  
For I cannot pretend to say. .

One wants society,  
Another, variety.  
Others a tranquil life;  
Some want food,  
Others, as good,  
Only want a wife.

But this poor little cat  
Only wanted a rat,  
To stuff out its own little maw;  
And it were as good,  
Some people had such food,  
To make them hold their jaw!

Betty can sometimes be persuaded to improvise a story or poem on a subject selected by another person. In such cases she dictates it, often more rapidly than one can take it down in longhand. One improvised poem of 16 lines she dictated to one of us in 2 minutes and 50 seconds. Twenty-one university students in an advanced class in versification were allowed fifteen minutes in which to write a poem on the same subject. This test was given by the regular instructor as a class exercise. The poems were then typed and given to 35 judges to be ranked in order of merit. In the combined ranking Betty's poem held sixth place. Besides Betty's, we reproduce the best and the poorest written by the university students.

#### A TALE OF OTHER BLACKBIRDS THAN THE ONES IN THE KING'S PIE

(Betty. Written at age 8)

Four and twenty blackbirds  
 Seated on a roof  
 One did peck his neighbor  
 His neighbor grunted, "OOf",  
 The neighbor also, he got mad  
 And pecked another bird,  
 His neighbor also wasn't glad  
 And said a naughty word.  
 This neighbor pecked another yet,  
 The neighbor vowed he'd not forget;  
 took vengeance on another fowl  
 And raised a long and doleful howl.  
 Then that one turned without a word  
 And pecked the very lastest bird,—  
 So all the birds got in a quarrel  
 Because of one ungodly fowl.

Moral: If you do good, the good will be mirrored in those around you. Also, if you be bad, the bad will be imitated by those around you.

#### BLACKBIRDS

(Rated poorest among the verses by university students)

I sat on a window sill,  
 The sun was shining bright;  
 A row of blackbirds chattering, fill  
 My mind with thots all right.  
 But to tell what I have in mind,  
 Is harder by far, than to guess  
 What the twitter of those birds mean,  
 As they spatter their words about.

I gaze on their pointed faces,  
And see by twirls and grimaces,  
The creatures are poking at us,  
We who think we are smart,  
At our expense and display,  
Fun of the grandest sort.  
One calls us worms, creatures that live in the clay,  
Another like grass dried up like hay.

### BLACKBIRDS

(Rated best among the verses by university students)

Out there on the roof in the sunshine  
Twelve blackbirds are stopping to rest.  
Their merry chirping and chatter  
Perchance, speak of spring and a nest.

We sit at our desks near the window,  
And painfully, patiently try  
To keep our eyes fixed on a lesson—  
Our minds on some old "Poly Sci."

Oh, why cannot we, like the blackbirds,  
Sit out in the bright morning sun  
And muse on the coming of summer  
When all of our lessons are done?

But we cannot fly like the blackbirds  
Far into the deep, azure sky—  
Instead,—we may soar after knowledge  
And mount after learning so high.

The majority of Betty's poems have been composed in less than ten minutes. The sight of a rainbow or sunset, or robin, perchance some experience while at play, or a suggestion from her reading, brings a flash of inspiration which sends her post haste to her typewriter. Recently, however, she has come to prefer dictation, because it leaves her freer to think. Hardly a word is changed during dictation and few afterward. Few false starts are made. At times she dictates more rapidly than anyone can write in longhand. While dictating her voice is low and her words run into each other so that they are hard to catch. Meanwhile she walks, skips or dances about over the room. Her stories are composed much less rapidly than her poems, perhaps at an average rate of 300 or 400 words in a half hour.

Betty's prose compositions have not been rated, but on the whole they seem to be about as remarkable as her poems. Both stories and poems show a prolific imagination, a lively sense of humor, and keen powers of observation. Her diction is usually good, but her sentence structure often shows carelessness and in poetry her meter sometimes limps. We repro-

duce a brief prose poem, a few lines of description from a story entitled "The Honeymoon," a fairy story, and one from a series of nature stories entitled "Stories the Alphabet Told" (one for each letter of the alphabet, written as a Christmas present to us).

#### FAIRY DEFINITION.

(Written at age 7 years, 11 months)

Fairies are the fancies of an imaginative brain,  
Which wearying of earthly realities aspires to  
create beings living only in thought  
endowing the spirits thus created  
with all genius for giving  
HAPPINESS.

P. S. I wonder why grown-ups scoff at faries  
And believe that angels have wings.  
I don't think one is any harder to  
Believe than the other,  
Do you?

#### PARAGRAPH FROM "THE HONEYMOON."

(Written at age 7 years, 10 months)

"They stopped a moment to watch the sunshine glistening on the stream and to watch the suckers, trout and minnows darting through the water. Once they caught sight of an eel. The forest was so interesting that day. At last after gazing a long time at the silvery fish they again entered into the wilderness. A little bunny scurried across the path, a robin chirruped, a little squirrel came so close they were almost afraid of him, a cow mooed afar off, a pair of grouse scurried across the road even as the rabbit had done. A keen looking fox slunk along to the east. Once they heard a bear growl. Once a gentle eyed fawn rushed close past them. A pair of quail with frightened calls rose from some brush and flew over their heads to the other side. Once they came across the nest of some baby goldfinches."

#### WHY THE DAISY HAS A GOLDEN HEART AND ROSY PETALS.

(Written at 7 years, 2 months)

ONCE the Daisy had no heart. Not even a sign of one. One day as she was loudly bewailing her fate like this—Ah me, Would that I had even a tiny stone to take the place of a heart. Would that—At that moment the Sun came by. What are you alasing about little one? he said with a jolly laugh. Alas the daisy replied, I am weeping because I have no heart. Ho, Ho, laughed the Sun. "Is that all"? THAT is easily remedied. You know the sun is all made of gold, so he pulled out some of his hair and fashioned it into the loveliest little golden heart you ever saw and laid it in the breast of the daisy. There little one, said the jolly sun. There take that and whenever you'r glad that you have a heart think of your old friend, the Sun, and he sailed away. (But the dear little daisy was modest and so timid and shy that what do you suppose she did,—She folded her petals tight around her golden heart. But in the morning she grew a little bolder and opened her pretty petals and all the flowers said, how wonderful the daisy has a golden heart.

End of part I

## PART II WHY THE DAISY HAS ROSY PETALS.

Little Daisy stood in all her beauty and the sun came by, there was a little butterfly on a sunbeam—snow-white with spots of gold like the daisy herself—So she called out—

"Little butterfly dearest come down and talk with me. But the butterfly did not hear. She called a second time—Then the butterfly heard and foulding his wings dropped right down on the daisy's petals. Oh you beautiful thing he said, you beautiful thing. And the Daisy blushed for joy—and ever since then the daisy has her rosy tinge.

Moral. the things we hope for may come to pass.

## STORK

(Written at age 8 years, 10 months)

S stands for Stork who brings babies 'tis said,  
But before any see him, he's back safe in bed.

Sallie Stork and Sammy Stork were heart-broken. They had been married some weeks ago, they had built a nest, but Sallie had laid no eggs as yet. "Oh dear," they bewailed in concert, "What will become of us if we have no children? Oh dear, oh dear, we must have some children, we must. But at last one day, Oh joy, could it possibly be? There in the nest lay a big, white egg. What should she do! Oh they were in paroxysms of delight.

They waited a long while and happy day, who came out of the egg but little Susan. Oh happiness! "Mamma," she inquired, in a sweet stork lisp, "Where am I? What's happened to me? The last thing I remember is good St. Nicholas taking me to one of the angels who put me in a tight white crib. What happened then, mamma, won't you tell me?"

"Well," her mother answered, "I don't exactly know, but it appears to me that when the angels put you into the egg, that's what it seems to me the crib was,—they brought you down and put you in our nest."

"Well, that's nice," said Susan, I'm sure I couldn't have found a better mother in heaven. You know there aren't generally any storks up there."

"Aren't there?" her mother answered, "Well I'm glad of that for perhaps if there had been many, you wouldn't have come down to us, you darling." "Indeed," chirped in Father Sammy, "I don't believe they could have brought us a finer young Stork.

So the days passed happily along and Susan grew tall and big and soon learned to fly.

At length one day she suggested to Sammy and Sally, "Why can't we take a trip up to Heaven now, I'm sure my wings are strong enough, why not?"

"Very well," Father Sammy answered, and Mother Sally agreed. "If you want to we will do it. When shall we start?" He enquired of her.

"Why not start to-day?" suggested their neighbor, Susie, brightly, who had come in for a little chat.

"Why not?" answered Susan, who was no longer little, "So here we go," and they soared high up in the sky, Susie included.

Finally they got there. All at once Susan gave a cry of joy—"St. Nicholas," she cried, and flew rapidly to-ward him. He was a jolly roly-poly fat man. "Why if this isn't Susan!" he cried gladly.

"My but you've been away a long time. But we are in great trouble," he went on, "I really don't know what to do. The angels are all busy taking down babies or watching over others and there are four human babies who have just arrived and have to be taken to their mothers. What shall we do? You know we're all occupied. The worst is that one can't take them alone, or I might do it, but they live in different parts of the world, one in Chicago, one in Hong-Kong, one in Amsterdam and one in Petrograd; and they all have to be delivered to-day C. O. D. The charge is 10 Lbs. of happiness C. O. D. Except the mother in China, she only pays five because they don't love girl babies in China as much as the boys. But as it's the only Chinese vociferous here, it will have to go, I suppose."

"I'll take the Chicago one," offered Susan.

"That's nice," cried old Santa. "Very obliging of you I'm sure. The address is, 1589 Lover Lane."

"Don't count me out of the game," intercepted Susan. "I'll pack the squaller of China."

"Well that address," said old Santa, "is XYZ Ching Chong Joy Path."

"I'm sure I'll help too," remarked Mother Sally. "Give me the little Dutchman. I've heard that little Dutch babies aren't as noisy as the others."

"Oh yes they are," St. Nicholas laughingly assured her, "But if you wish you can take charge of that fellow," and she flew off to get the baby.

"Well I guess I'll take a chance if the others' do," Father Sam decided "And as the baby Russian is the only one left, I'll take that."

"All right, that's on Lovesky Lanesky. Deliver it to the Countess Cuddlesky," and stooping with difficulty he pasted on the youngster's stomach, a label bearing, "Glass,—With care," and Sammy flew off.

"There" said old Santa drawing a big sigh of relief, "I do wonder when they'll come back."

In a little while Susan came back. "Gracious goodness, that woman gave me 20 Lbs and didn't ask for the change," she remarked as she settled down beside him.

Old Santa laughed with satisfaction. "Well she was certainly generous she must be a millionaire at the joy bank."

In another minute back came Susie. "You said that woman should only be charged 5L. but she handed over 25L. Let that 'not loving' stuff be forgotten, why I bet if I had waited long enough she'd have given me five million."

And then came Sally. "Ten pounds did you say? She handed over ten hundred it seemed to me. Gracious they must have a million millions of happiness by the way they spend it."

At last came Father. "Aren't the Russians generous? You said ten pounds, didn't you? They gave ten thousand and offered to give more if I'd wait a little."

"I've enjoyed the trip so much," Susan timidly hinted;

"That I think we might" (from Susie)

"Be the Saint's (this from Sallie)

"Messengers." (this from Father Stork.)

"That's right," and they looked at each other in satisfaction.

"That's right" Old St. Nicholas approved, "We'll take you into our service," And ever since the storks have had no occasion to be idle

Moral. The only blissful joy, you know,

Is, in making some other one's happiness grow.

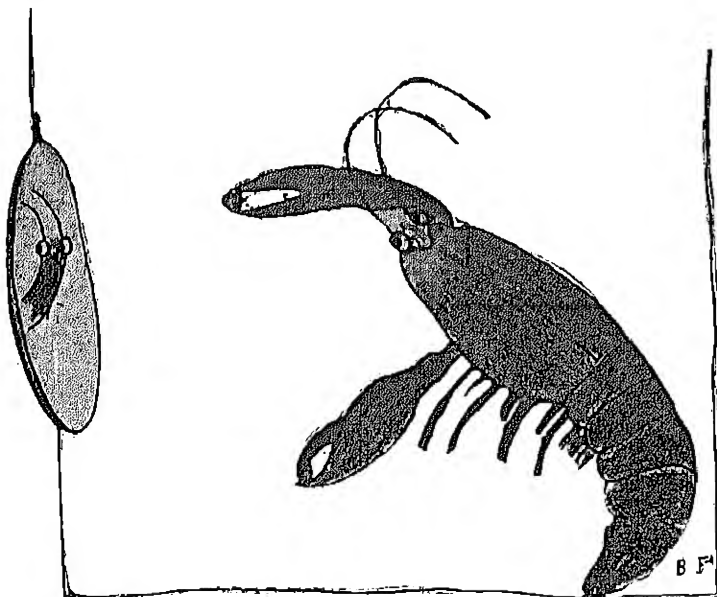


PRELIMINARY REPORT ON A GIFTED JUVENILE AUTHOR 175

Stork. Family—Ciconiidae, are large and white except the extremities of the wing tips, which are black. Their food consists of frogs, birds, reptils and insects. Storks build their nests on chimneys, trees or high rocks. The eggs are white. Its common name in Dutch means, 'Bringer of Good.'

Many of her poems and stories she has illustrated with colored crayon or pencil drawings. Some of these give evidence of considerable talent. Her pictures are never copied directly. If she needs a model (as for some animal picture) she finds as many different pictures as she can and then makes her own, using the best features from each model, together with such changes and additions as suit her purpose. She draws very carefully, making many corrections; then, when she is at last satisfied, she traces her final outline with carbon paper in order to secure a neat copy, and colors it with crayon.

Very frequently she appends to story or poem a "moral," well pointed and aptly expressed. One not knowing her might attribute this moralizing tendency to priggishness, but



LOBSTER.

it is really due to her playful enjoyment of the intellectual gymnastics involved in generalizing a situation or incident.

Although Betty's compositions show a wide range of interests, there is a marked preference for themes relating to nature. Classification of 86 poems gave the following result: nature poems (descriptive and appreciative), 44; reflective poems, 11; prose poems, 8; homely poems, or poems of everyday affairs, 7; love poems, 6; occasional poems (Christmas, election-day, etc.), 4; plays and extravaganzas (one a play of 1,500 words), 3; unclassified, 1. Of 65 prose compositions, 31 are nature stories of the *pourquoi* type ("Why the bunny has a white cotton tail," "Why the daisy has a golden heart and rosy petals," "Why the owl loves the moonlight," "Why the thistle has thorns," etc.), 25 are animal stories (in a collection entitled "Stories the Alphabet Told"), 6 are fairy tales, 1 a narrative story of 1,500 words (entitled "The Honey-moon") and 1 a detective story of 2,500 words.

It is not possible at present to make any quantitative statement regarding the degree of originality which characterizes Betty's work. Of course no one would expect to find the compositions of a child of eight years entirely lacking in derivative character. Not infrequently one comes upon familiar themes and even familiar phraseology. There are so many common elements in the poetry and in the stories of all ages that the most creative of writers can hardly hope to strike many new notes. Absolutely original thoughts are perhaps even more rare in literature than they are in science, because literature is older and more universal than science. It would be psychologically absurd not to recognize it as inevitable that such an impressionable and sensitive child as Betty should draw many of her inspirations from her reading. Occasionally, though very rarely, her voracious memory seems to have swallowed entire sentences, but there is no reason to believe that her borrowings have exceeded the limits of unconscious absorption.

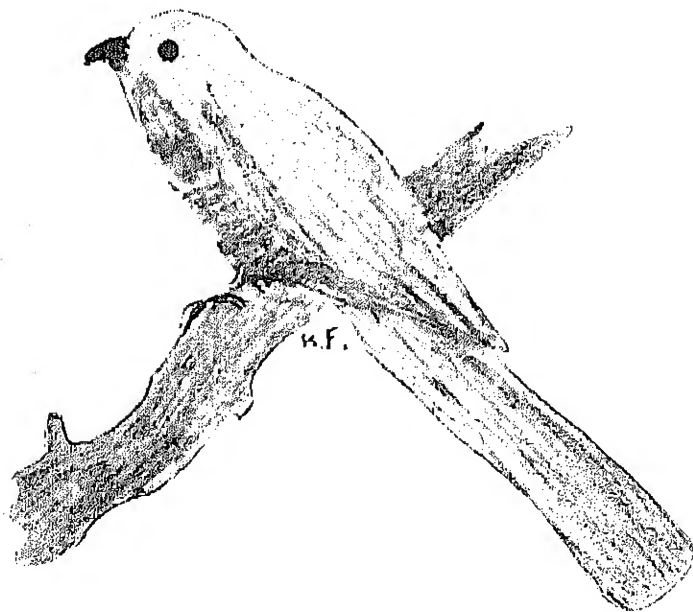
Except for slight myopia, Betty has no known physical defects, and has never been seriously ill. She sleeps soundly from 10 to 11 hours a day. At the age of 8 years and 2 months her height corresponded to the standard for 10½ years, her weight to the standard for 9½ years, her grip to the standard for 10 years, and her lung capacity to the standard for 8½ years. It is perhaps because of her unusual activity that she prefers to play with boys rather than with girls. In doll play and other ordinary occupations of girlhood

# PRELIMINARY REPORT ON A GIFTED JUVENILE AUTHOR 177

she has shown only moderate interest. Her ambition is to be a physician and a writer.

Ratings of Betty on 46 physical, intellectual, emotional, volitional, social and moral traits were secured from the mother and from a school teacher who had known the child intimately since babyhood. The ratings were made on a scale of 1 to 5, 1 being defined as "very superior," 2 as "superior," 3 as "average," 4 as "inferior," and 5 as "very inferior." The average of the mother's ratings was 1.43; of the teacher's ratings, 1.47. The mother's ratings included only two as low as "average" (those on intellectual modesty and communicativeness). The only rating by the teacher as low as "average" was that for beauty.

Our own rating on the traits which we have had a chance to observe would not differ greatly from those of the mother



CUCKOO.

or teacher, although they might be somewhat lower on intellectual modesty and social adaptability. In neither of these traits, however, does there seem to be any marked inferiority, if one makes due allowance for the difficulties which are inevitable in such an extraordinary case of intellectual precocity. It is probably this, chiefly, which explains the fact that she prefers the companionship of children who are either several years older or several years younger than herself. The mother's habit of restricting her associations with children of her own age has doubtless tended to confirm this natural but not altogether wholesome preference. There is no evidence that when she does play with other children there is any marked difficulty of social adjustment. Doubtless Betty knows that she has talents which other children do not have. She is too intelligent to be kept in ignorance of the fact. But if she is conscious of her superiority she gives little evidence of it. The mother has tried to imbue her with the idea that if she has been endowed with any unusual gift it is her duty to use it to make the world happier and better. "Mother," she said one day, "I am not proud because I can write verses; they just come themselves. But if I could only learn to control my temper, then I would be proud."

Betty's education has been managed by her mother, whose guiding principle has been the conviction that a child's confidence in its own ability to think and do should be sedulously cultivated. She believes that the abilities of children often atrophy for lack of encouragement or because of the destructive criticism of their elders. When Betty does a thing well she is commended. Spontaneity is cultivated. Her literary ability cannot be regarded as an artificial product of special training. It was her mother's ambition that she should become a musician. When it became evident that the child's talents lay in another direction she was encouraged to follow her own bent.

## ON THE NEED FOR CAUTION IN ESTABLISHING RACE NORMS

By ADA HART ARLITT, Bryn Mawr College\*

In recent years there has been a tendency to establish race norms on the basis of mental tests given to representatives of various races who have emigrated to the United States. The results of tests made of groups of children of foreign born parents have been stated and comparisons made with the distribution of intelligence in groups of children of American born parents. So far as we have been able to ascertain except in the case of negro children<sup>1</sup> no attempt has been made to check up the influence of factors, other than that of race, which might quite as well account for the differences which it has been assumed were traceable to race alone. Particularly is this true in regard to the part played by social status. Much has been written about the effect of social status on the median intelligence in any group of American children, little or nothing has been said as to how much this may account for the low or high median score or median intelligence quotient in groups of children of other races. For instance Terman<sup>2</sup> speaking of the influence of race on Intelligence Quotient cites the median I.Q. for Italian children as being 84 or 16 points below that usually given for American children.

It was with a view to determining the relative influence of the two factors, race and social status, that this investigation was begun.

The tests used were the Stanford Revision of the Binet tests. These were given by the author and the four graduate students whose names appear as collaborators. All of the students were experienced testers. In order to insure uniformity of grading all tests were marked in consultation with the author.

The children 343 in all were taken from the primary grades

---

\* (With the assistance of C. Baechle, I. Neterer, K. Smith, and C. Barnes.)

<sup>1</sup> Sunne Dagny. A Comparative Study of White and Negro Children. *Jour. of App. Psychol.*, 1917, p. 71.

<sup>2</sup> Terman. *Intelligence of School Children*, p. 56. (Houghton Mifflin Co., 1919.)

in a single school district. Of these 191 were children of native born white parents, 87 were Italians and 71 were negroes. All of the Italians spoke English without difficulty.

The native born white children were divided on the basis of social status into five groups with reference to the occupation of the father. The separation into groups by occupation followed Taussig's division into the five non-competing groups, i.e., (1) Professional classes; (2) Semi-professional and higher business; (3) Skilled; (4) Semi-skilled; (5) Unskilled. The home conditions followed closely the division by occupation. The last two groups, children of unskilled and semi-skilled workers, contained too few children to be treated separately. They were therefore combined. As a matter of convenience these children will be called throughout the "Inferior Group," the children whose parents fall into the skilled class will be called "Average," those of the semi-professional and higher, business class "Superior" and those of the professional class "Very Superior."

Of the Inferior group there were 43, of the Average group 76, of the Superior group 48 and of the Very Superior group 24.

TABLE I

Social.....	50-	60-	70-	80-	90-	110-	120-	130-1	Above	Total
Status.....	59.9	69.9	79.9	89.9	109.9	119.9	129.9	139.9	140	..
V. Sup.....	0	0	1	0	6	4	6	3	4	24
Sup.....	0	0	0	0	12	17	14	4	1	48
Av.....	0	1	3	2	41	16	10	2	1	76
Inf.....	1	3	5	9	23	1	0	1	0	43

The effect of Social Status on the distribution of Intelligence Quotients.

The outstanding characteristics of each group can best be seen from Table I. Over 40% of the Inferior social status group have I.Q.'s below 90 and only 4.6% have I.Q.'s above 110. There are only two children of very superior mentality in this group. In the Average group 50% are of average mentality and over 25% are above average i.e., have I.Q.'s above 110. The curve for the Superior group shows it to be as much better than the Average group as that was superior to the Inferior social status group. The median I.Q. is 118.7 or 10.2 points above that of the Average group. Seventy-five per cent of these children have I.Q.'s above 110 and none have I.Q.'s below 90. A curve for the Very Superior group

would have two modes, one at 90-110 the other at 120-129. Seventy per cent of these children have I.Q.'s above 110 and of these 76% have I.Q.'s above 120. The difference between the Superior and Very Superior social status groups lies not in the proportion of the children who reach 110 or above but in the fact that a larger proportion of those whose I.Q.'s pass 110 reach 120 or above.

It is interesting to note that very few children of superior mentality were found in the Inferior social status group and that only one child of inferior mentality was found in the Superior and Very Superior social status groups.

The median I.Q.'s for the four groups were respectively 92, 107, 118.7 and 125.9 or a difference of 33.9 points between children of Inferior and Very Superior social status of the same race and attending the same grades in the same school.

We are well aware, as has been stated, that previous writers<sup>3</sup> have emphasized the part played by social status. It is here cited chiefly because the variation due to social status is wider than that usually reported and serves to demonstrate all the more clearly the part that this factor may play in determining the distribution of intelligence in any non-selected group.

The effect of social status on the native born white group having been determined the total group was divided on the basis of race alone. It is a significant fact that there were too few children in the Italian and negro groups of even average social status to enable us to study the combined effect of race and social status in these groups.

The median I.Q. for the native white group was 106.5,<sup>4</sup> for the Italian group 85 and for the negro group<sup>5</sup> 83.4. The native born white group had a median I.Q. 21.5 points above that of the Italian and 23.1 points above that of the negro group.

<sup>3</sup>Wientrob, J. and R. *The Influence of Environment on Mental Ability*, *Jour. of Educ. Psychol.*, 1912, pp. 577-586. Terman, L. M. *The Stanford Revision of the Binet-Simon Scale*, Chap. V. (Warwick and York, 1917.) Terman, *Intelligence of School Children*, pp. 188-189. (Houghton Mifflin Co., 1919.)

<sup>4</sup>The high median I.Q. is probably due to the fact that most of these children were between the ages of five and nine. The range of the I.Q.'s made by the middle fifty per cent of the children tested by Terman was from 97 to 111 in five and six year olds combined, and from 95 to 111 in seven and eight year olds combined. The average child might therefore be expected to make a higher I.Q. at these ages than at later ages as for instance thirteen to fourteen where the limits of the quotient including the middle fifty per cent are from 90 to 105. (Data cited from "Stanford Revision of the Binet Tests," page 40.)

<sup>5</sup>The negroes were of mixed blood. It was impossible to determine the proportion of white blood in the case of any of these children.

This disparity is great and if it had been due to race alone would have demonstrated beyond a doubt the superiority of the native white group, but 37% of the native white group came from families of Superior and Very Superior social status whereas 93% of the negro and 90% of the Italian group came from families of Inferior and Very Inferior social status.

The Italian and negro children were then compared with the group of native born white children of Inferior and Very Inferior social status. The median I.Q. for the native white group was then 92 or 8.6 points above that of the negro and 7 points above that of the Italian group.

This difference is not due to a larger proportion of children of superior or very superior mentality in the native

TABLE II

	50- 59.9	60- 69.9	70- 79.9	80- 89.9	90- 109.9	110- 119.9	120- 129.9	130- 139.9	Above 140	Total ..
Native born White.....	1	3	5	9	23	1	0	1	0	43
Italian.....	3	11	19	17	26	4	1	0	0	81
Negro.....	1	7	16	20	25	1	1	0	0	71

The effect of Race on the distribution of Intelligence Quotients.

white group. There are five such children in the Italian group, two in the negro and one in the native born white group. It is due to the proportion of children in the negro and Italian groups whose I.Q.'s are below 80. In the native born white group 20.9% have I.Q.'s below 80, in the negro group 33.8% have I.Q.'s below 80 and in the Italian group 41.2% fall below 80. As all of these children live in the same section of the same community and attend the same grades in the same schools this difference both as to the median I.Q. and as to proportion of dull children seems to be racial. But the difference in median I.Q. which is due to race alone is in this case at most only 8.6 points whereas the difference between children of the same race but of Inferior and Very Superior social status may amount to 33.9 points. It is apparent that such differences as we have between the negro and Italian children and between these

<sup>a</sup> These three groups were originally the same size namely, 29 in each. Other cases have been added to bring the groups up to their present size with a consequent change in the median I.Q. of only .5 of a point. The median I.Q., therefore, seems to represent a difference which is not due to the difference in the size of the groups.



and children of native born white parents are not nearly so striking as the difference between children of the same race but of different social status. Of the two factors social status seems to play the more important part. To such an extent is this true that it would seem to indicate that there is more likeness between children of the same social status but different race than between children of the same race but of different social status.

While the number of children tested is too small to permit us to make statements which may be universally applicable our results seem to warrant the following conclusions:

1. There is a difference in the distribution of intelligence in negro and Italian groups as compared with groups of children of native born white parents which seems to be due to race alone.

2. There is a marked difference in the distribution of intelligence in groups of children of the same race but different social status. This disparity seems to be greater than the disparity between children of the same social status but different race.

3. Race norms which do not take the social status factor into account are apt to be to that extent invalid.

## THE PREDICTIVE VALUE OF SHORT INTELLIGENCE TESTS

By C. F. HANSEN and M. J. REAM, Carnegie Institute of Technology

Just how long an intelligence test should be to insure reliability is a question that has been frequently raised. The use of intelligence tests in industry has often been hindered by the length of time required for giving them. The Army Alpha test requires forty-five minutes, the Thorndike Comprehensive Tests approximately three hours. An investigation has been made with an intelligence test devised by the Bureau of Personnel Research of the Carnegie Institute of Technology, to determine the effect of greatly shortening the time of the test upon: (1) standing as determined by the full time test and (2) correlation with a practical criterion. The results show that an intelligence test like Personnel Bureau Test VI can be considerably shortened with practically no loss of reliability or predictive value. Accordingly an adaptation of the test to the time available—whether five minutes or a half hour—has been secured.

Personnel Bureau Test VI is a simple modification of the Army Alpha intelligence test.<sup>1</sup> Six parts of the Alpha have been "scrambled" to make a continuous test, the oral directions and the common sense questions only being omitted. The remaining items were arranged in "spiral omnibus" form, i. e., they become progressively harder as the test advances. The whole series is continuous, since complete directions covering all the classes of items appear at the beginning, after which the subject works steadily without interruption for the entire allotted time. The test in this form has been given to more than a thousand salesmen and office workers and is in regular use in the employment offices of the Westinghouse Electric and Manufacturing Co., the W. T. Grant Co., the Phoenix Mutual Life Insurance Co. and others.

In determining the reliability and predictive value of short intelligence tests, experiments have been made with two groups of students in the School of Life Insurance Salesmanship at Carnegie Institute of Technology. The groups consisted almost entirely of men between the ages of twenty-one and

---

<sup>1</sup> A description of this test has been previously published. Bingham, W. V. Some Group Tests. *Psychol. Bull.*, Feb., 1920, p. 57.

forty-five, all of whom were eager to meet the requirements for success in this school.

The testing of these two groups followed, with one modification, the usual procedure of allowing four minutes for reading the directions and illustrations and twenty-five minutes for the test itself. In response to a signal given by the experimenter at intervals of five minutes each, the subjects made a check mark opposite the item upon which they were then working. The papers therefore furnished five scores, as follows:

- (1) Score for the first five minutes.
- (2) Score for the first ten minutes.
- (3) Score for the first fifteen minutes.
- (4) Score for the first twenty minutes.
- (5) Total score on the complete test of twenty-five minutes.

In the first experiment Test VI Form A was given to the first group, forty-eight students, in October, 1920. The relations between partial scores and total scores are shown in the following table of correlations:

	Total Score	P.E.
Score for 5 minutes.....	.88	$\pm .02$
Score for 10 minutes.....	.92	$\pm .02$
Score for 15 minutes.....	.96	$\pm .01$
Score for 20 minutes.....	.97	$\pm .01$

In the second experiment Test VI Form B was given to the second group, eighty students, in January, 1921. The correlations between partial scores and total scores were as follows:

	Total Score	P.E.
Score for 5 minutes.....	.89	$\pm .01$
Score for 10 minutes.....	.93	$\pm .01$
Score for 15 minutes.....	.97	$\pm .01$
Score for 20 minutes.....	.98	$\pm .01$

The scatter diagrams showed that the close relationship between partial and total scores held true of all the individuals, not a single widely diverging case being found.

To test further the reliability of the shortened tests the partial scores were correlated with an objective criterion, viz., the grades obtained in all school courses taken by the students. According to the system in use at the Carnegie Institute of Technology, the school grades were translated into quality points. A grade of A was given four points for quality, a grade of B three points, a grade of C two points, a grade of D one point, grades of E and F no points for quality. As each person in the insurance school took exactly the same

courses the total quality points were directly comparable for all the students.

The correlations of partial scores with quality points earned in the school courses were as follows:

	Quality Points	P.E.
Score for 5 minutes.....	.53	$\pm .07$
Score for 10 minutes.....	.54	$\pm .07$
Score for 15 minutes.....	.63	$\pm .06$
Score for 20 minutes.....	.63	$\pm .06$
Total Score.....	.62	$\pm .06$

The results show that an intelligence test of only five minutes is fairly proportional to the complete twenty-five minute test, a test of ten minutes more closely proportional. By the use of the regression equations scores in the short time tests can be converted into approximate full time scores and thus be compared with established norms. In order to get the intelligence score for the complete twenty-five minute test:

Multiply the score made in *five* minutes by 2.03 and add 42; or

Multiply the score made in *ten* minutes by 1.38 and add 28; or

Multiply the score made in *fifteen* minutes by 1.15 and add 13; or

Multiply the score made in *twenty* minutes by 1.02 and add 11.

This means that the intelligence test can be made short or long according to the time at the disposal of the examiner.

For practical purposes in predicting school success, the *fifteen minute test is just as satisfactory and reliable as the longer test.* Though the reliability increases slightly with the longer periods, the fast worker and the slow one alike maintain their gait throughout the test. Differences in ability appear in the first five minutes of a test, and these differences in ability are roughly proportional throughout. To continue an intelligence test of this character beyond a point of proven reliability is merely to duplicate results already secured.

## BOOK REVIEWS

*Lehrbuch der experimentellen Psychologie.* By J. FRÖBES. S. J. Herdersche Verlagshandlung, Freiburg. Bd. I, 1916, 605 pp.; Bd. II, 1920, 704 pp.

These two volumes constitute a remarkably complete compilation of psychological investigations culled from a wide and varied domain and done with a thoroughness typically Teutonic. The author's purpose was to bring together the material of experimental psychology as it exists to-day in a manner at once comprehensible for introductory work and suitable as a point of departure for advanced specialization, somewhat after the fashion of Tigerstedt's *Lehrbuch der Physiologie*. A glance at the wealth of matter in the book shows that the author has been far more inclusive than the title might indicate, for branches of psychology other than experimental are treated at considerable length.

In the first volume Fröbes deals with the simpler psychological processes. The order of treatment is conventional. The first section, following an introductory chapter on the object and methods of empirical psychology, discusses the nature and conditions of sensation in general, specific nerve energies, correspondence between sensation and stimulus, psychophysical parallelism, and leads directly to a survey of the special senses in section two. Here we find detailed treatment of the sensory phenomena of each sense department. The chapter on vision, *e. g.*, begins with the structure of the eye and takes up in turn the psychological arrangement of visual sensations, the relation of color sensations to the stimulus, color mixture, adaptation, indirect vision, contrast, after-images, the temporal course of visual sensations, color blindness, and concludes with a statement of visual theories. No less complete, although not so lengthy, are the chapters on audition, taste and smell, tactual sensations, kinaesthetic and static sensations, and organics. A chapter on the simple feelings closes section two. Section three, which is devoted to perception, deals with simple and complex ideas (*Vorstellungen*), tonal fusion, consonance and dissonance, visual and tactual space perceptions, and perceptions of time and movement. An excellent presentation of the subject of psychophysics is given in the next section. The development and practice of psychophysics from Fechner down through Wirth and Urban—but not Thomson!—with a confessed predilection for the views of G. E. Müller, are outlined without undue sidesteps into points of controversy. Fröbes' own investigations in this department are referred to quite modestly in passing. A good chapter on correlation brings the section to an end. An hundred page section on the association of ideas completes the first volume.

The second volume leaves the field of experimental psychology for those wider domains which serve as supports for language, aesthetics, criminology, sociology, education, medicine, law, *etc.* In the first volume the author was well at home in his own special branch of study; lacking this intimate familiarity with certain of the subject matter of the second volume he had the sagacity to seek expert advice from *Fachmänner*. An opening section treats of disturbances of association and cerebral localization; the so-called higher mental processes,

attention, the self, memory, reasoning, imagination, come in for detailed discussion; over an hundred pages are devoted to the emotions, aesthetic feelings, mimicry, and physiognomy; well-known authorities are quoted at length in chapters on the will, morals, customs, religion, personality, mental development, and social psychology; and the volume is brought to a close with some 125 pages on psychopathology. A valuable survey for those wishing an orientation into the topics treated!

This work of Fröbes is unquestionably the most useful source of reference for general psychology which has appeared to date. It would seem, however, that a Handbook of Psychology would be a more fitting title than a Textbook of Experimental Psychology. It does not attempt to be systematic and to that extent would be dangerous in the hands of those beginners who have no guidance from a particular point of view. But to those students who have gotten grounded somewhere in psychology and want reliable reference to less familiar topics, this work will come as a tremendous contribution. The author has selected his literature wisely and widely (some 400 references to writings in English, 230 of which are by Americans!) and has not littered the pages with inverted commas, colons, and semicolons, which verbatim quotations necessitate, but has given in clear, concise form (sometimes painfully concise: *ebenso* Witasek) an epitome of the views of authorities cited.

July, 1921, Clark University

C. C. PRATT

*Fundamentals of High School Mathematics. A Textbook Designed to Follow Arithmetic.* By HAROLD ORBWAY RUGG and JOHN ROSCOE CLARK, both of the Lincoln School, Teachers College, Columbia University. Cloth xv+ 368 pages. Illustrated. Price \$1.80. Also Answer Book, 16 cents. Yonkers-on-Hudson, New York: World Book Company.

Every teacher or school administrator who is interested in improving the educational value of the first-year high school mathematics course should examine this new book. Unlike many recent texts in algebra this one has a real excuse for being.

The authors have assembled in this ninth grade course the most important mathematical tools and notions which all children should know. The excessive manipulation of symbolisms of formal algebra has been replaced by material which is, both from the mathematical and social point of view, of far greater value. Specifically, graphs, methods of direct and indirect measurement, the properties of the right triangle, and a comprehensive treatment of the concept of relationship between changing quantities (*i. e.* functionality) have replaced the elaborate treatment of factoring, fractions, and operations with long polynomials.

The selection of material for this course is quite in agreement with the recommendations of the National Committee on Mathematical Requirements. The authors are to be congratulated upon having sensed fully the best thought of the day in the reorganization of first-year high school mathematics courses.

Among the unique impressions made upon the reviewer, the following stand out most prominently:

- (1) The careful explanations and development of new processes;
- (2) The wholesome omission of formal material;
- (3) The excellent presentation of word-problems;

- (4) The unique organization of special products and factoring;
- (5) The "timed practice-exercises" for developing skill in essential tool processes; and
- (6) The emphasis upon the notion of relationship between variable quantities.

*Manhood of Humanity; The Science and Art of Human Engineering.*  
By ALFRED KORZYNSKI. E. P. Dutton & Co., New York, 1921,  
p. 264.

This is a book of very wide scope embracing all human problems. It is a study of man and the problems of humanity from a scientific, especially a mathematical, point of view. The author makes a plea for the establishment of a "Science and Art of Human Engineering" dealing with human problems not by means of a study of metaphysical essences, but of actual facts, and reduces these facts to mathematical laws. The author attempts to explain social problems by an application of mathematical reasoning and mathematical law. "Engineering is the co-ordinated sum-total of human knowledge gathered through the ages, with mathematics as its chief instrument and guide."

In his first chapter, the author deals with a new concept of life and makes use of a very interesting mathematical analogy which although in the opinion of the reviewer, somewhat overdrawn, nevertheless serves to bring out in a very striking manner how the progress of the social sciences has lagged behind that made by the natural sciences. The second chapter takes up The Childhood of Humanity and shows how metaphysical speculation, selfish political philosophies and sectarian opinions have hindered the progress of the social sciences and thus prolonged the period of humanity's childhood right up to the present day. In the next chapter the author gives an unique classification of the types of life which points out vividly the relation of man to other classes of life and shows how the great importance of man lies in his "time-binding capacity." This classification is fundamental in the development of the author's theory. The entire next chapter is devoted to a systematic inquiry into the nature of man remarkably free from metaphysics and dogma in an attempt to arrive at a scientific understanding of the function and sphere of man. This is perhaps the most important chapter in this important book and the concept here developed by the author is of far-reaching importance in the scientific solution of humanity's problems.

Chapter V Wealth and Chapter VI Capitalistic Era outline the method of approach of the new science of Human Engineering to these fundamental economic problems. The author points out that economic ideas "however false and harmful, are protected alike by habit and by the inborn conservatism of many minds" and how these old ideas hinder the natural growth of our social sciences. These economic problems are attacked in a clear, logical way and an attempt is made to get at fundamental economic concepts, reducing them to mathematical law. This is a practical application of the author's theory to the field of economics. Chapter VII deals with the Survival of the Fittest in a useful way and points out the difference in the way survival works in animals and in men. Chapter VIII, The Elements of Power, endeavors to explain power by mathematical and mechanical methods and takes up the German philosophy of power that had so much to do in bringing on the World War. In Chapter

IX, *Manhood of Humanity*, the author shows that a scientific basis for the study of the social sciences will initiate the *Manhood of Humanity*, when the social sciences will prevent the energies of man from being wasted by ignorance and selfishness, by competitive conflict characteristic of beasts, but will show that these energies are more than sufficient to produce a high order of increasing prosperity everywhere throughout the world. The author gives a necessarily vague but suggestive outline of a plan for science and art of Human Engineering to guide the large affairs of the new state of humanity's manhood. This state is to be controlled by scientists and not by petty politicians. "The world will have uninterrupted, peaceful progress when and only when the so-called social 'sciences'—the life-regulating 'sciences' of ethics, law, philosophy, economics, religion, politics, and government are technologized; when and only when they are made genuinely scientific in spirit and method."

The book also contains Appendix I which takes up Mathematics and Time-Binding; Appendix II, Biology and Time-Binding; and Appendix III, Engineering and Time-Binding giving a choice bibliography for each appendix.

*Manhood of Humanity* is a valuable attempt to a true conception of what human beings really are—an attempt to initiate the scientific study of man—a science and art that will know how best to direct the energies of man to the advancement of human welfare. It is written in a clear, logical, stimulating style and the theory advanced by the author will probably arouse much controversy. It is an attack of a very old subject from a new point of view and is a truly remarkable contribution toward the scientific study of humanity and should command the attention of all interested in humanity's problems.

MAX MEENES

*Methods and Results of Testing School Children.* By EVELYN DEWEY, EMILY CHILD and BEARDSLEY RUMF. E. P. Dutton & Co., New York, 1920, p. 176.

This book is the result of a Psychological Survey, conducted in the New York City Public Schools in February, 1915. "The immediate purpose of the investigation was to obtain norms for a series of tests for New York public school children in the poorer and more congested portions of the city, and a basis for further study of the value of mental tests for improved school procedure. It was hoped that the results would justify the establishment of a school clinic for normal children to assist the teacher in meeting the problems of individual pupils." Accordingly, a series of tests was devised to measure numerically five differences of performances between children of the same chronological age and it was aimed to accomplish this by presenting situations requiring mental adjustments of the type that a child uses in adapting himself to his environment. The authors confess that although improvements could have been made at every stage of the investigation, they limited the changes almost wholly to the elimination of tests that were unsatisfactory in some phase of administration in order to avoid an impracticable repetition of the tests. It is perhaps significant to note that in giving the test the authors adopted a method in which the maximum amount of rigidity and uniformity were prescribed, under the conviction that a really quantitative measure of the child's adjustment was their goal and would justify the limits in flexibility which this method would imply.



The authors based their investigation on results obtained from Jewish subjects and aimed further to collect a parallel set of results in schools where children of other nationalities predominated. Pupils between the ages of eight and thirteen were tested. A large variety of "tests" were given to the children individually and under standard conditions. An effort was made to avoid the child's receiving the impression that he was being classified according to his performance in the tests or that he was undergoing a school examination. The testing was never begun until the child felt at home with the tester and interested in doing as he was asked. His mistakes were never corrected, the tester maintaining an encouraging and satisfied manner no matter what the response." This was not a group intelligence test, but an individual performance scale containing a number of separate special tests such as cancellation, tapping, tests of manual dexterity, cart construction and some tests from the Yerkes-Bridges and from the Binet scales. The authors give a description of these various tests and the directions for giving them, together with the methods used for working up the results. The results are treated statistically in the many tables and charts contained in the book. An added feature is a number of charts showing sex differences brought out by the tests used in this investigation. A home investigation of the children tested was also made to provide data for an extensive social study of the homes from which the children came and the social data thus obtained is contained in tabular form. To complete the survey, all the children tested were given a thorough physical examination and this book contains tables of norms obtained in the physical examination of these children.

The investigation is rather thorough as far as it goes and the work was necessarily slow and painstaking, but such an investigation must be pushed further to include larger numbers of children of all nationalities and all environmental conditions so that satisfactory norms may be established and Psychological Clinics formed in connection with the school departments of various cities.

MAX MEENES,

## NOTES

### INDUSTRIAL RESEARCH LABORATORIES

Research facilities and the development activities of American industries are to be described in the forthcoming revision of *Bulletin of the National Research Council* Number 2, "Research laboratories in industrial establishments of the United States of America." Only 300 such laboratories were listed in the first edition but it is hoped that several hundred new names will appear in the revision and that a more nearly complete reference list will thus become available. The general demand for the first edition of the Bulletin shows the wide interest in this subject, and the importance of having every laboratory which devotes even a portion of its time to research properly listed.

The Council requests information from directors of research who have not already supplied it. The following data are wanted: Name and address of firm and address of laboratory; name of director of research; number on laboratory staff (classified as chemists, engineers, bacteriologists, etc.); approximate proportion of time spent on research; chief lines of research; unusual features of equipment; research laboratory space; date of organization of research laboratory and annual expenditure for research. Confidential information is not desired.

It is also requested that librarians in the service of the industries please bring this notice to the attention of the proper officials in their organizations.

This material should be furnished as promptly as possible to the Research Information Service, National Research Council, 1701 Massachusetts Avenue, Washington, D. C.

### REPORT OF PSYCHOLOGICAL WORK IN THE UNITED STATES ARMY

A complete report of the history, methods and results of psychological examining in the United States Army has been recently published in the *Memoirs of the National Academy of Sciences*, Volume 15, 1921. The report is edited by Lieut. Col. Robert M. Yerkes, Chief of the Division of Psychology, as an official document for the Surgeon General of the Army. It consists of three parts bound in a single volume. Part I, presenting the official history of the development of the service and its activities during the war, is supplemented by reproductions of all of the printed materials devised and used in conducting psychological examinations. Part II includes a complete account of the preparation of methods, their characteristics, and their evaluations as practical procedures. In Part III the results of examining are summarized.

The entire report may be obtained from the Superintendent of Documents, Government Printing Office, Washington, D. C., at \$1.75 per copy. It appears in quarto size under the title "Psychological Examining in the United States Army," and includes VI+890 pages.

The following books and pamphlets have been received:

Charles S. Myers. *Mind and Work*. G. P. Putnam's Sons, New York and London. 1921. pp. 175.

- Stewart Paton. *Human Behavior*. Charles Scribner's Sons. New York, 1921. pp. 465.
- Rugg and Clark. *Fundamentals of High School Mathematics*. World Book Co. Yonkers-on-Hudson, N. Y. 1921. pp. 368.
- Harl Roy Douglass. *The Derivation and Standardization of a Series of Diagnostic Tests for the Fundamentals of First Year Algebra*. University of Oregon Publication, Eugene, Ore. Vol. 1, No. 3. April, 1921.
- Arnold Gesell. *Exceptional Children and Public School Policy*. Yale University Press, New Haven, Conn. 1921.
- Clara H. Town. *Analytic Study of a Group of Five and Six-Year-Old Children*. University of Iowa Studies, Iowa City, Ia. Vol. I, No. 4. May 1, 1921.
- Bird T. Baldwin. *The Physical Growth of Children from Birth to Maturity*. University of Iowa Studies. Vol. I, No. 1. June 1, 1921.
- Bassett and Porteus. *Sex Differences in Porteus Maze Test Performance*. Reprint from the *Training School Bulletin*. Vineland, N. J. No. 22. Dec., 1920.
- Bureau of Juvenile Research. *Review of the Work, 1918-1920*. Columbus, Ohio. Publication No. 19. Feb., 1921.



# THE JOURNAL OF APPLIED PSYCHOLOGY

VOL. V

SEPTEMBER, 1921

No. 3

## THE RESULTS OF REPEATED MENTAL RE-EXAMINATIONS OF 639 FEEBLE-MINDED OVER A PERIOD OF TEN YEARS

By F. KUHLMANN

Early in 1912 the writer adopted the intelligence quotient method of grading intelligence. This was before Stern's monograph in which this method is proposed appeared.<sup>1</sup> Results from the examination of 1,300 feeble-minded of all ages quoted at that time seemed to indicate that this ratio of mental age to age remained roughly constant through the increasing ages of the feeble-minded. It was pointed out, however, that this constancy required that normal or average mental development progress at a uniform rate from year to year, as measured in absolute units of measurement, an assumption contrary to general observation and theory. In 1911 a ten-year program was begun of re-examining all inmates of the Minnesota School for Feeble-Minded at regular intervals of two years, excluding cases over twenty years of age, all epileptics, and others in whom some special trait interfered with getting a reliable mental age. The object of this study was to test out thoroughly the value of the I. Q. method of classification, and to determine the traits of the mental growth curve.

The examinations involve 639 cases from the grade of idiocy to nearly average normal intelligence, and from one to twenty years of age. Each case was examined from two to five times. About a third of the examinations were made by the writer. The others were made by Maud A. Merrill, Dr. Frances Lowell, Katharine B. Graves, and Rose Anderson, past and present research assistants. All had had ex-

---

<sup>1</sup> See Kuhlmann, F. "Degree of mental deficiency in children as expressed by the relation of age to mental age." *Journ. Psycho-Asthenics*, June, 1913.

tensive preparation in psychology and thorough training in the use of the tests. Variability of results due to different examiners was probably a negligible factor.

Three disturbing factors were encountered. The first concerned the cases examined. A case once examined did not always remain in the institution for future re-examinations. Some were absent temporarily at the time a re-examination was due. Others were removed permanently through death, discharge, or transference to other institutions. The second factor was the possible selective influence on grade of intelligence of the institution in admitting and discharging cases at different ages. It could not be assumed that cases admitted or discharged at different ages were of the same grade of intelligence. The third factor was the varying accuracy of the tests at different age-levels, and of the different revisions of the B.-S. scale that were used. The first examinations were made in 1910 with the original 1908 scale.<sup>2</sup> From 1912 to 1917 my 1912 revision of the tests was used.<sup>3</sup> Since the beginning of 1917 my last revision was used. The first two gave on the whole quite the same results, the 1912 revision being concerned chiefly with standardization of procedure and extending the scale downwards below the age of three.<sup>4</sup> The last revision corrected the general tendency of the earlier scales of giving too high mental ages towards the lower end, and too low mental ages towards the upper end, and increased the number of tests to eight for each age group.

The first and second of these disturbing factors could be met satisfactorily. The influence of the third could be eliminated only in part. This will be considered in connection with the statement and discussion of the data.

#### *Average Yearly Increase in Mental Age*

The raw data gives a varying number of cases whose first examinations came at different ages, from one to eighteen years. Except for the disturbing factors just mentioned, the average mental growth curve for these inmates of the institution could be at once determined by computing the average mental age at each age of all cases examined and re-examined.

<sup>2</sup> See Kuhlmann, F. "Binet and Simon's system for measuring the intelligence of children." Journ. Psycho-Asthenics, 1911.

<sup>3</sup> See Kuhlmann, F. "A revision of the Binet-Simon system for measuring the intelligence of children." Journ. Psycho-Asthenics, Monograph Supplements, Sept., 1912.

<sup>4</sup> See Kuhlmann, F. "Some results of examining a thousand public school children with a revision of the Binet-Simon tests." Journ. Psycho-Asthenics, March and June, 1914.

This procedure, as a matter of fact, gives a growth curve surprisingly close to what is obtained when the disturbing factors are eliminated. The number of cases at each age is large enough to give a fairly smooth curve in spite of the fact that at each age some of the previous cases have dropped out and other new ones have been added. The selective influence of the institution, as shown also by special methods of treating the results to determine it, is small, and negligible for most parts of the growth curve. The mental ages at each age, since the cases include all grades, include a considerable range, and thus partly eliminate or smooth out the varying errors in the mental ages due to inaccuracies in the scale at different levels.

But these relationships change when we attempt to determine the mental growth curves separately for the different grades of intelligence. The difficulty is increased by the fact that our definition of grade of intelligence in terms of the intelligence quotient involves us in assumptions the correctness of which constitutes our chief problem. We may attribute a certain range of I. Q.'s to a given grade of intelligence for any particular age, but to use the same range for all ages for this grade of intelligence assumes, of course, that the I. Q. remains constant for all ages.

Table I gives the average mental ages for five successive years of each group of cases whose first examinations occurred at the same age. This is given separately for the four grades of intelligence termed idiots, imbeciles, morons, and borderline cases. The manner of deriving these figures needs to be carefully noted in detail. It was done through the I. Q.'s, as that procedure had certain advantages over that of dealing directly with the mental ages. The first step was to correct the individual mental ages found for errors due to the general tendency of the scale to measure too high or too low at different levels. This was done by subtracting from each individual mental age the amount the scale was known to measure too high on the average at this point. There were practically no mental ages at levels high enough to need additions as corrections. These corrections could not be made safely for mental ages below four, and could therefore not be extended to low grade imbeciles, idiots, or to the younger morons and borderline cases. Our norms for these lower mental ages are not sufficiently well established for any scale to allow of any definite conclusion as to what minor degree of error they may still contain. The data used in making these corrections were the smoothed figures on the average mental ages found with my 1912 revision on 1,000 public

school children, referred to above, and unpublished data for my last revision.

The second step was to compute the I. Q.'s for the ages between two successive examinations by taking the average of the one preceding and the one following. It was then easier and more accurate to compute the average mental age at a given age from the average I. Q. than it would have been to work with the mental ages directly. The examinations did not, of course, occur at the exact ages of seven, eight, and so on, and to have averaged the mental ages of all cases classified as examined at seven, for example, might have been more or less misrepresentative, as the exact average age of these cases could vary from six and a half to seven and a half years. All age classifications were made on the basis of the nearest birthday. The error was eliminated by computing the average mental age from the average I. Q., using the ages as exactly seven, eight, and so on. Results of fourth and fifth examinations of a case were thrown out in order to put all cases on the same basis of a five year period with three actual examinations at two year intervals. Data for ages below seven are not given in this table because the mental ages could not be corrected for the general tendency to error in the scale, in the case of the morons, for ages below this.

The third step aimed to eliminate any selective influence of the institution there might have been on grades of intelligence at different ages, and also the difficulty arising from classifying the cases into the four grades of intelligence at different ages on the basis of the I. Q. This procedure was as follows. (a) Computing the average I. Q. for morons, for example, examined first at the age of seven, and again at nine and eleven. (b) Computing the average I. Q. at eight for those examined first at eight, and again at ten and twelve. (c) Eliminating from this second group of cases low or high grades cases so that the average I. Q. of the remaining cases was approximately the same as the computed average I. Q. at eight for the cases examined first at seven. If there was a general tendency of the I. Q. to change from seven to eight, this procedure insured having the same true grade of intelligence for the group examined first at eight as for the group examined first at seven. On account of the small number of cases at times for such a group, this equating of the I. Q.'s could be done only approximately in many instances. The remaining step was taken to make such exact equating unnecessary. (d) The average I. Q.'s at ten and twelve were



TABLE I

No.	IDIOTS																			
	7	8	9	10	11	12	13	14	15	16	17	18	19	20						
6		1.60	1.88	2.17	2.44	2.72														
4			1.88	2.17	2.48	2.53	2.54													
5				2.17	2.36	2.55	2.71	2.86												
11					2.43	2.65	2.88	3.04	3.19											
2						2.61	2.78	2.93	2.97	2.81										
3							2.73	2.91	3.08	3.04	3.00									
2								2.93	3.12	3.10	3.08	3.07								
3									3.09	3.04	2.99	2.87	2.75							
										3.00	3.03	2.97	2.75							
No.	IMBECILES																			
	7	8	9	10	11	12	13	14	15	16	17	18	19	20						
8	2.80	3.17	3.53	3.91	4.28															
14		3.17	3.53	3.88	4.14	4.38														
16			3.53	3.77	3.99	4.28	4.44													
21				3.85	3.87	4.40	4.63	4.94												
24					4.07	4.21	4.30	4.38	4.44											
18						4.32	4.44	4.52	4.75	4.66										
25							4.46	4.71	4.94	4.84	4.73									
11								4.64	4.88	4.79	4.73	4.67								
12									4.75	4.74	4.74	4.76	4.78							
14										4.76	4.71	4.65	4.79	4.92						
10											4.73	4.33	3.93	3.85						
7												4.60	4.68	4.75						
													4.55	4.51						

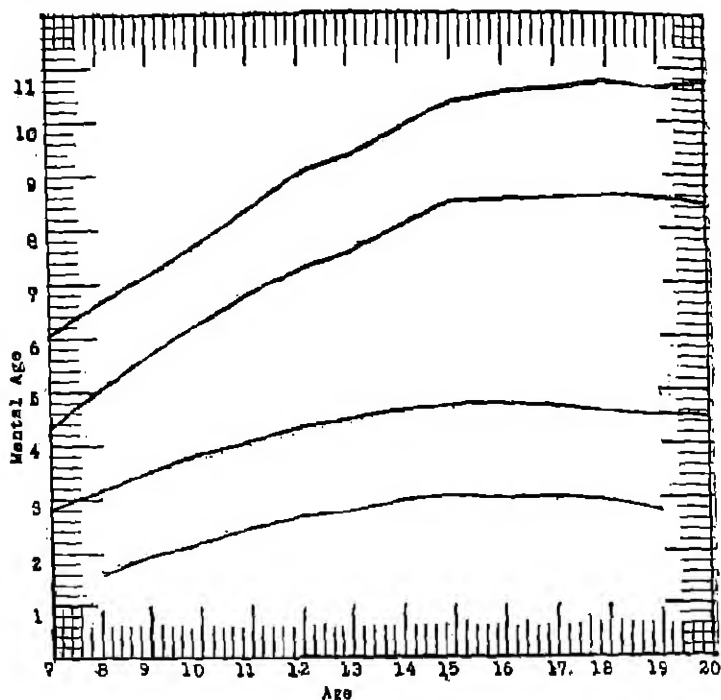


found for the selected cases examined first at eight, and the intermediate I. Q.'s computed as before for nine and eleven. The average I. Q. at eight for this second group was then placed at exactly the same as the computed I. Q. at eight for the first group examined first at seven. The change in the I. Q. from eight to nine for this second group was then added to or subtracted from the I. Q. at eight to get the I. Q. at nine. Likewise, the I. Q.'s at ten, eleven, and twelve were found for the second group by adding or subtracting the amount of change that had occurred since the previous year. This procedure was repeated for each group examined first at nine, ten, and so on. Thus the course of these derived I. Q.'s for each group examined first at a given age remained exactly the same in form as that of the averages of the actual I. Q.'s found in the examinations, but it might lie at a slightly different level, in the majority of instances in a level not over five points in I. Q. higher or lower.

In this table the number of cases examined for the first time at a given age is indicated in the first column of figures on the left. The last figure under any age gives the average mental age for that age. The total number of cases on which each average is based may be computed by adding up the figures on the left. The average mental growth curve from seven to twenty years for each of the several grades of intelligence is then given in these last figures for each age. The following graphs are based on the average mental ages in Table I.

The main features of these growth curves present but little that is not now pretty well known. They show that the feeble-minded do develop mentally, and at a rate in proportion to the grade of intelligence. This was still a disputed point at the time this study was begun. This rate of development decreases with age for all grades. The rate of this decrease cannot, of course, be shown with this data, as we do not know how much mental growth in terms of absolute units the mental year represents at different age levels. In the graphs each addition of a year in mental age is represented as equal to preceding additions of a year, but even so the curves decline with increasing age. Could they be plotted in terms of absolute units of growth they would decline very much more. This will be made more obvious below, in discussing the intelligence quotient.

*Age of Cessation of Mental Age Increase.* Space will not be taken here to review in detail previous studies and dis-



cussions on this matter. It will be remembered that Stern<sup>5</sup> assumed that the age of cessation of mental development changed markedly with grade of intelligence, the idiot grade ceasing developing relatively very soon after birth. Others since have accepted this assumption. Recently Doll<sup>6</sup> has claimed that his results on the re-examination of feeble-minded verified this assumption of earlier arrest for the lower grades. Grading his cases on the basis of the final mental ages they attain in reaching maturity, he summarizes as follows.<sup>7</sup> I

Final mental age....	1	2	3	4	5	6	7	8	9	10
Age of arrest.....	?	?	7	11	11	12	12	12	15	15

find myself unable to agree with Doll that his own results in this study will bear the interpretation given in these figures.

<sup>5</sup> See Stern, W. "The psychological methods of testing intelligence." Trans. by G. M. Whipple, 1914.

<sup>6</sup> See Doll, E. A. "The growth of intelligence." Princeton contributions to psychology, 1920.

<sup>7</sup> See page 76.

The Goddard revision of the B.-S. tests, which was used in this study, is not adequate for the determination of mental ages much below five, as it does not include standardized tests for ages below three. Doll's average growth curve for cases with a final mental age between 6 and 7 rises markedly beyond the age of 12. The same is true of his next curve, for cases with a final mental age between 7 and 8. His curve for cases with a final mental age between 9 and 10, and the next for cases with a final mental age between 10 and 11 show a still more obvious rise beyond the age of 15.<sup>8</sup> In fact, my own interpretation of Doll's results would be that they do not disagree seriously with my present results, based on a larger number of cases. In considering my own table and graphs, it should be noted that the different grades are defined in terms of the I. Q. at the age of seven,—eight for the idiot grade. At this age, cases with I. Q.'s from 0 to 24 were classed as idiots, cases with I. Q.'s from 25 to 49 as imbeciles, 50 to 74 as morons, and over 74 as borderline. The upper and lower limits of these ranges of I. Q.'s for these different grades then decreased with age, as explained above. It is seen that the average growth curves extending from the age of seven to twenty all rise to the age of fifteen at least. They indicate that idiots, as here defined, develop to the age of fifteen, inclusive, imbeciles to fifteen or sixteen, morons to about seventeen, and borderline cases to about eighteen. In other words, all grades develop much longer than Stern and others have supposed, and the difference for different grades is not nearly as great as has been assumed. These facts are shown even more strikingly in the I. Q. curves below than in the present mental age data. Since the mental age of cases of fifteen years or over was always divided by fifteen to get the I. Q., the I. Q. rises beyond this age, when the mental age continues to increase.

*Frequency of Yearly Gain and Loss in Mental Age.* In considering this question the results for all grades of cases will be combined. It will not be necessary to take into account the three disturbing factors noted above. Not correcting the mental ages for errors in the scale will also enable us to include all the results for mental ages below seven. Since the mental age increase for the two-year interval between two successive examinations is nearly always much less than two years, and since any two successive mental ages will both be affected in the same direction through the error in the scale, the amount of loss or gain in mental age between two

---

<sup>8</sup> See page 75.

TABLE III

Age	Number Gaining	Number Stationary	Number Losing	Per cent Losing
2	2	0	0	0
3	6	0	0	0
4	13	1	1	7
5	19	0	0	0
6	17	1	1	5
7	30	0	0	0
8	41	0	1	2
9	57	2	2	3
10	52	2	2	4
11	47	4	13	20
12	35	0	13	27
13	42	5	5	10
14	33	3	8	18
15	16	5	14	40
16	16	9	11	30
17	11	6	12	38
18	16	6	18	45

compared, 91 per cent of the lower grade cases that lose occur after the age of ten, while 94 per cent of the higher grade cases that lose occur after the age of ten. It seems, therefore, that age is also a factor determining deterioration quite independently of grade.

#### *Constancy of the Intelligence Quotient*

The method of determining the continuous course of the I. Q. from seven to twenty years from results that are limited to four year periods of examinations and re-examinations for any given case has already been explained in presenting the mental age growth curves. Table IV gives the I. Q.'s that correspond to the mental ages given in Table I. In plotting the average, continuous I. Q. curve from these figures, however, it becomes necessary to take into account the variations between the different groups of cases whose first examinations occurred at given ages. The irregularities in the average curves are thereby at once explained. Thus, in the moron grade especially, the cases examined first at seven happen to be improving cases, and their I. Q.'s cause a rise in the average I. Q. curve from seven to nine.

These results determine certain tendencies of the I. Q. It decreases with increasing age for feeble-minded. That decrease is larger for the high grade than for the low grade cases. Nearly every group examined first at a given age shows both these facts. The averages, given in the last figure under each age, showing the continuous course of the I. Q. to age twenty, reveal them more strikingly. That the I. Q. of the feeble-minded would decrease with age was anticipated, but that it would decrease more for the higher grade than for the lower grade cases was contrary to expectation. I shall show later that both these tendencies of the I. Q. follow mathematically from the fact of a certain type of decreasing rate of growth with increasing age of average normal children, when that growth is measured in absolute units of measurement instead of by the variable unit given in mental ages.

The above figures do not decide the question as to whether the decrease in I. Q. takes place at a uniform rate for all ages for a given grade, or whether it decreases more rapidly at some ages than at others. The rise in the I. Q. beyond the age of fifteen is due, of course, to the fact already noted that the mental age continues to increase beyond this age, while fifteen was the highest age by which the mental age was divided to get the I. Q. Not considering ages beyond fifteen it seems from these figures that there are no very marked changes in the rate of decrease of the I. Q. for different ages, and that the course of the I. Q. is roughly a straight line. A different method of treating the results, given below, however, shows some unquestionable age effect. Taking that part of the average I. Q. curve for each grade that is not obviously disturbed by accidentally varying groups, gives the following average yearly decline in the I. Q. for each of the four grades.

Borderline.....	2.19 points <sup>9</sup>
Morons.....	1.21 "
Imbeciles.....	1.04 "
Idiots.....	.37 "

<sup>9</sup> In studying the results to determine this average yearly decrease in the I. Q. several different methods were used for eliminating the disturbing factors discussed at the beginning of this article. These methods gave somewhat different figures, but all showed that the I. Q. decreased more with the higher grades than with the lower grades. Even when no attempt was made to eliminate these factors, and when the same range of I. Q.'s was attributed to a grade at all ages, and the average I. Q. computed at each age as found in the raw results these figures were, for borderline 2.2 points yearly loss, for morons 1.3, for imbeciles .5, and a negligible change for idiots.

TABLE IV

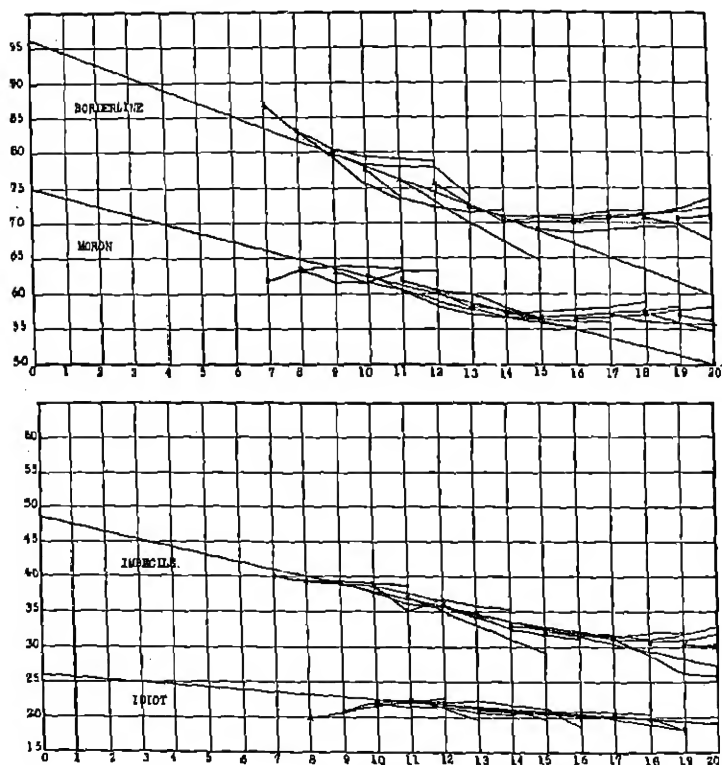
No.	Idiots																												
	7	8	9	10	11	12	13	14	15	16	17	18	19	20															
6		20.00	20.84	21.61	22.17	22.67																							
4			20.84	21.70	22.56	21.06	19.56																						
5				21.69	21.49	21.29	20.85	20.40																					
11					22.07	22.12	22.16	21.71	21.25																				
2						21.79	21.37	20.95	19.83	18.71																			
3							20.99	20.76	20.52	20.27	20.02																		
2								20.96	20.81	20.66	20.56	20.46																	
3									20.60	20.27	19.93	19.13	18.33																
										19.98	20.17	19.80	18.33																
No.	Imbeciles																												
	7	8	9	10	11	12	13	14	15	16	17	18	19	20															
8	40.00	39.63	39.26	39.10	38.93																								
14		39.63	39.20	38.77	37.63	36.49																							
16			39.23	37.73	36.23	35.70	34.17																						
21				38.53	35.20	36.63	35.97	35.30																					
24					37.00	35.05	33.05	31.28	29.46																				
18						35.97	34.14	32.30	31.69	31.08																			
25							34.34	33.64	32.94	32.24	31.54																		
11								33.13	32.54	31.95	31.54	31.13																	
12									31.66	31.64	31.62	31.75	31.87																
14										31.73	31.38	31.02	31.91	32.80															
10											31.52	28.87	26.22	25.67															
7												30.69	31.19	31.68															
													30.30	30.05															



No.	MORONS																
	7	8	9	10	11	12	13	14	15	16	17	18	19	20			
11	61.73	62.80	63.88	63.55	63.60												
10		62.80	62.67	62.53	63.25	63.05											
31			63.23	61.68	60.40	59.02	57.59										
24				62.60	60.47	58.36	57.20	56.70									
25					61.93	60.31	60.00	58.08	56.37								
21						60.19	58.06	56.84	56.03	55.21							
17							58.21	57.34	56.51	56.23	56.02						
24								57.24	57.51	57.79	58.20	58.66					
17									56.61	56.98	57.25	57.56	58.20				
17										56.55	56.86	57.17	57.81	58.20			
14											57.08	56.09	55.93	55.69			
13												57.37	56.09	54.83			
													57.01	56.24			

No.	BORDERLINE																
	7	8	9	10	11	12	13	14	15	16	17	18	19	20			
6	86.67	82.83	79.33	75.67	73.33												
9		82.83	80.50	79.50	79.16	78.94											
10			79.92	78.02	78.06	77.92	74.22										
10				77.73	73.63	72.40	71.60	71.80									
5					76.05	73.05	70.05	67.45	64.85								
5						75.58	72.78	70.18	70.18	70.18							
12							72.16	70.99	70.91	70.60	71.02						
10								70.11	70.81	70.90	71.71	71.82					
5									69.19	68.79	69.19	69.39	69.39				
4										70.12	70.87	71.02	72.02	73.52			
5											70.70	71.30	71.90	72.50			
8												70.88	69.75	67.58			
													70.76	71.20			

The straight lines in the following graphs approximate very closely this average rate of decline, assuming that it is the same for all ages for a given grade. It will be seen in a moment that this assumption cannot be entirely correct. In these graphs the course of the I. Q. for each group of cases examined first at a given age is plotted separately. The small crosses mark the course of the average I. Q. curve. It is seen that this average curve misrepresents the more probably true curve in several instances, most markedly for the beginning of the average curves for the idiot and moron grades.



*Relation of Rate of Decline of the Intelligence Quotient to Age.* It was seen already that the percentage of cases whose mental age decreases as they grow older increases with age. So far as this might be a general tendency for all cases this effect of age should be more marked for the I. Q., for the

I. Q. may decrease considerably with age without causing an actual loss in mental age. Table V gives the percentage of cases at each age whose I. Q. decreases. The results for all grades are combined, and include no eliminations or corrections for the disturbing factors considered in connection with some of the other questions already discussed. Correcting the mental ages for errors in the scale was found not to make any material difference. The I. Q.'s for intermediate years between two successive examinations of a case were computed and these computed I. Q.'s were used for the results in this table.

TABLE V

Age	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13	13-14	14-15
Number cases	15	42	58	71	122	151	186	233	215	163	182
% decreasing	20	19	45	49	60	71	76	78	75	67	66

These figures show an unquestionable age effect on the decrease in the I. Q. that is not revealed in the figures of the preceding table. The frequency of cases losing in I. Q. increases with age up to about eleven years, the age at which the frequency in mental age loss was found to increase rather suddenly. When the results are computed separately for each of the four grades each grade shows an age at which the frequency of loss in I. Q. is at its maximum and beyond which it declines again, just as for the results of all grades taken together. For idiots this maximum age is twelve, for imbeciles eleven, for morons ten, and for borderline cases seven and twelve. The average mental ages for these grades at these maximum ages are 2.6, 4.1, 6.3, and 6.1 and 9.1, respectively. This relation of frequency of I. Q. decrease to age is therefore not the effect of errors in the mental ages due to the scale of tests. If this were the case the maximum frequency in loss in I. Q. should have occurred approximately at the same mental age. When the mental ages are corrected for these errors, as could be done for mental ages of four and over, the number of cases left for each age becomes rather small for reliability. It makes no material difference for the borderline cases, but raises the maximum age for the morons from ten to thirteen. The correction should, of course, have made more difference for the borderline group than for the moron. We may conclude with safety only that the frequency of loss in I. Q. increases with age, irrespective of grade, up to a certain age. Whether or not this frequency declines beyond a certain maximum age is not so certain.

*Relative Frequency of Loss and Gain in Intelligence Quotient.* One of the chief questions concerning the I. Q. has been the possibility of predicting future and final mental ages on its basis for the individual case. If the present I. Q. has been determined for a case in a given single examination, how well can his mental age for any age in the future be predicted? Unfortunately this question has been confused with the question of the constancy of the I. Q., with which it has no necessary connection. The ability to make this prediction does not, of course, depend on its constancy but on the regularity of its change in successive examinations at different ages, if there is a change. The traits of the I. Q. established by the present results have complicated its use as a means of prediction. It decreases with age, and the amount of decrease for any given year is dependent on the two further factors of age and grade. Accurate prediction requires that we know not only the presence of these tendencies, but also their extent at each point. But, having discovered these tendencies, we can make corresponding allowances in predicting future mental ages, and thereby reduce a general tendency to error in prediction that would otherwise be present. The following frequency distribution results on the loss and gain in I. Q. gives a fair idea of the general reliability of the I. Q. when used for prediction without allowing for the tendencies to change that was found. Table VI gives the number and percentages of cases that lost or gained 1, 2, 3, 4, etc., points a year. In this table the results of all examinations are again included without eliminations or corrections. They are grouped according to the four grades, and irrespective of age. This, of course, introduces the error into the classification resulting from attributing the same range of I. Q.'s at all ages to a given grade. Since we now know the general tendencies of the I. Q. to change with age and grade the effect this may have on the distribution will be understood. The task of determining the true range of I. Q.'s for each grade at different ages, taking the decline with age of the I. Q. into account, is too complicated for satisfactory solution. Our previous task of determining what cases should be classed under each grade at different ages involved only getting approximately correct average I. Q.'s at each age, which was a relatively simple matter. It is evident that the present procedure cannot give a materially different distribution from what would result if the error in question were eliminated.

TABLE VI

		Over -10	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
Idiots. . . . .	No.								1	6	25	35	48
	%								.6	3.6	15.9	22.6	30.5
Imbeciles. . . . .	No.					2	5	6	23	44	178	235	179
	%					.2	.6	.6	2.7	5.4	20.6	27.2	20.6
Morons . . . . .	No.		3	3	11	3	15	49	71	119	133	149	100
	%		.3	.3	1.2	.3	1.5	5.0	7.3	12.2	13.6	15.3	10.2
Borderline . . . . .	No.	5	10	6		6	43	34	34	23	25	22	11
	%	1.8	3.6	2.2		2.2	15.6	12.3	12.3	8.3	9.0	8.0	4.0

TABLE VI—Continued

		+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	Av.
Idiots. . . . .	No.	19	9	9		3				2		-.06
	%	12.1	5.7	5.7		1.8				1.2		
Imbeciles. . . . .	No.	114	33	23	5	2		2				-.55
	%	13.2	3.8	2.7	.6	.2		.2				
Morons. . . . .	No.	95	75	65	34	13	4	10	10	2		-.72
	%	9.7	7.7	6.7	3.5	1.3	.4	1.0	1.0	.2		
Borderline. . . . .	No.	10	17	13	5	4	3	2				-2.7
	%	3.6	6.2	4.7	1.8	1.5	1.1	.7				

The following summaries from the figures in this table might serve as a rough guide in making predictions, which can then be much improved by taking into account the central tendencies of the I. Q. to decrease with age as found above.<sup>10</sup>

<sup>10</sup>I hope at some time later to discuss more fully the general problem of prediction, in which the present data will be treated specially with reference to this question.

IDIOTS			IMBECILES		
Per cent.	Points	Change	Per cent.	Points	Change
31	1		21	1	
65	2		61	2	
87	3		85	3	
96	4		94	4	
97	5		97	5	
99	6		98	6	

MORONS			BORDERLINE		
Per cent.	Points	Change	Per cent.	Points	Change
10	1		4	1	
35	2		16	2	
57	3		31	3	
75	4		44	4	
86	5		58	5	
93	6		72	6	
94	7		88	7	
96	8		91	8	
98	9		91	9	
			94	10	

These figures are to be read as follows. Of the idiots 31 per cent change less than 1 point in I. Q. a year; 65 per cent change less than 2 points in I. Q. a year, etc. It is seen from even this rough method of stating the results that much is left to be desired. Yet, we are far from justified in concluding, as is often done, that the I. Q. is of no value at all in making predictions. And further, if predictions are to be attempted at all, what other procedure or method have we that would be even approximately as reliable as predicting on the basis of the I. Q.? Aside from the question of prediction, the I. Q. remains the most accurate and convenient method available for expressing grade of intelligence at any given age of a case.

#### *The Normal Absolute Growth Curve*

It remains now to account for the traits of the I. Q. changes with age as found in the present results. The traits to be considered are, (1) that the I. Q. on the whole decreases with age; (2) that this decrease is largest for the highest grade cases studied, those with an I. Q. over .75, and becomes smaller the lower the grade; (3) that loss in I. Q. increases at least in some measure with age.

Considering the general principles of the B.-S. scale, the method of establishing norms and of deriving the mental age and I. Q., there seem to be three possible suppositions that might explain these general facts wholly or in part. The first

is that these traits of the I. Q. result directly from changes in the rate of growth, even when that rate of growth is measured in terms of absolute units. This would mean that the idiot child grows at a retarded rate that kept approximately the same relation to that of the average normal, losing only slightly in relative rate with age. As this retarded rate, however, approaches that of the average normal, giving us the higher grades of subnormals, this loss in relative rate increases, at least up to the grade called borderline in the present study. It is obvious that this loss in relative rate must then decrease again as we approach the average normal rate still closer, since at exactly average normal rate this loss becomes zero and the I. Q. remains constant at 1.00. The correctness of this first supposition is so entirely improbable on the face of it as to really need no further consideration.

A second supposition is that these traits of the I. Q., or at least the first two, may be the result of error in standardizing the scale of tests. It is assumed that the norms for the tests at each age are for non-selected children, representing true averages of all children at each age. This assumption is probably not entirely correct for any scale of tests yet devised. The public schools have always furnished the children in question, and it is recognized that the schools have a selective influence on the average intelligence of the children at each age or school grade. As we pass from younger to older, higher and higher grades of the subnormals are eliminated from the schools through failure to keep up, increasing the average level of intelligence of those that remain. The effect of this process of elimination on the I. Q. with a scale of tests standardized with such so-called non-selected school children would then be as follows: The I. Q. as found would decrease with age for all children of true average intelligence or less, since at each higher age the scale would measure a little bit more too low. It would not increase with age for children above average intelligence. The total decrease in I. Q. for a given period of years, let us say to sixteen, would be the larger the higher the grade of case in question, because the higher grade cases would attain a higher final mental age than the lower grade cases. To make this clear, let us suppose that our scale of tests were standardized in this way for the ages of one to sixteen. Let us assume a group of 100 truly non-selected children at birth, from which one child with a true I. Q. of .70, will drop out each year through some such selection as the schools exert on children in the schools. If norms for tests were established with the

remaining of these 100 children at each age, the I. Q.'s obtained with such tests would be .3 point too low at the age of one, .6 too low at two, .9 too low at three, and so on, to 4.8 points too low at sixteen. With such tests, the average child would lose 4.8 points in I. Q. in sixteen years; the child with a true I. Q. of .75 and attaining a final mental age of twelve at the age of sixteen would lose 3.6 points; the child with a true I. Q. of .50 and with a mental age of eight at sixteen would lose 2.4 points, and, so on. Now it is more than probable that the schools exert this selective influence, but there is no data available at present to indicate exactly how much effect this may have had on our norms for our tests. It is possible, therefore, to explain the first two traits of the I. Q. of the feeble-minded, at least in part, on the basis of this resulting error in the scale of tests. I believe that the present results are in some measure due to this factor. The third trait of the I. Q. however, cannot be explained on this ground, nor would it explain an increase in I. Q. with age for children above average intelligence.

The third supposition that may be made to explain the three traits of the I. Q. is that for the average child the yearly increments in mental growth as measured in terms of absolute units decrease each year, giving the general type of growth curve usually assumed to be correct. I shall attempt to show now that this assumption explains not only the three traits of the I. Q. found in the present results, but also the increase in I. Q. found by others for children above average intelligence. This will also suggest certain other traits of the growth curves of subnormals and of the I. Q. that the preceding analysis of the results has not revealed.

In a recent discussion Freeman<sup>11</sup> presents two types of normal growth curves either of which it is claimed would result in the constancy of the I. Q. for cases developing at a subnormal rate. The first is the logarithmic curve,  $y = \log. x$ . The second is the straight line, assuming a constant rate of development from year to year. For each normal or median curve he plots a mental age curve of a hypothetical case so that at any age the mental age of this case is .66 of the mental age of the median at the same age. That is, the growth curve for the hypothetical case is so plotted in both instances that the I. Q. *will* remain constant at .66. Freeman does not discuss the fact that if the ratio between the median rate of growth and the subnormal rate, measured in absolute

<sup>11</sup> The Interpretation and Application of the Intelligence Quotient. J. Ed. Psychol., Jan., 1921.



units, is assumed to remain constant the logarithmic growth curve cannot give constant I. Q.'s for subnormals. In the case of the straight line type of growth curve for the median, any subnormal developing at a slower than median rate and with a constant I. Q. will also maintain a constant ratio to the median rate of growth, when growth is measured in terms of absolute units, which in this case would be a year of median growth. Now for the logarithmic growth curve this is not true at all. Here these ratios, determined from the y values of the subnormal and median curves at each age, increase with age while the I. Q. remains constant. Thus, for the illustration used by Freeman, the figures run as follows:

Age.....	3	6	9	15
I. Q.....	.66	.66	.66	.66
Ratio.....	.61	.78	.80	.85

Vice versa, if the rate of growth of a subnormal maintains a constant ratio to the median rate, the I. Q. cannot remain constant, but will decrease with age. Since we have found as a matter of fact that the I. Q. does decrease with age, we must look for the explanation in a median growth curve that will give this decrease in the manner found.

It would be a relatively simple matter to determine the exact nature of the median growth curve if we knew the exact course of the I. Q. from birth to mental maturity, and if we could assume that the rate of development for any grade of intelligence above or below median maintained a constant ratio to the median, when rate is measured in terms of absolute units. We could then construct an absolute median growth curve that would fit the course of the I. Q. changes. We can determine the whole course of the I. Q. from birth to maturity for the grades of cases studied only by assuming that it declines at a uniform rate for all ages. It was seen above that this assumption is probably not very far wrong, the general course of the I. Q. being approximately a straight line, with a small tendency for the rate of decline to increase with age. We may tentatively accept also the assumption that the rate of growth of any grade maintains a constant ratio to the median rate, and then construct an absolute median growth curve that will fit the rate of decline of the I. Q. of one of the four grades of our cases, and with that median growth curve determine the course of the I. Q. for other grades of intelligence. This will show the following. (1) That the absolute median growth rate decreases each year, giving a type of curve similar to the logarithmic, but with the yearly increment in growth decreasing at a much slower rate

than for the curve  $y=\log. x$ ; (2) that this type of growth curve accounts for all the traits of the I. Q. that were established in the above results; (3) that the I. Q. for grades of intelligence above the median will increase with age instead of decrease.

I will take the moron grade, assume that its I. Q. declines at all ages at the uniform rate of 1.21 points a year, which was the average rate of decline found for morons. This gives this grade an average I. Q. of .74 at birth, and of .5585 at the age of fifteen. This gives the following I. Q.'s at the different ages of 0 to 15, with the corresponding mental ages.

TABLE VII

Age	0	1	2	3	4	5	6	7
I. Q....	.74	.7279	.7158	.7037	.6916	.6795	.6674	.6553
M. A....		.73	1.43	2.11	2.77	3.40	4.00	4.59

TABLE VII—Continued

Age	8	9	10	11	12	13	14	15
I. Q..	.6432	.6311	.6190	.6069	.5948	.5827	.5706	.5585
M. A.	5.15	5.68	6.19	6.68	7.14	7.58	7.99	8.38

To construct an absolute median growth curve that will give these I. Q.'s we may set the total growth during the first year from birth to age one arbitrarily at 100 units. Then

$$1 + \frac{.74y-100}{x} = .7158; \text{ also, } y=100+x, \text{ when } x \text{ equals the}$$

increment in growth from one to two years, and  $y$  equals the total number of units of growth at age two. Here .74 $y$  is the total number of absolute units of growth at age two for this grade of case, and 100 is the total number of units of growth at age one for the median.  $1 + \frac{.74y-100}{x}$  is then the mental age of this case, which, divided by 2, equals the I. Q. .7158. Solving for  $x$  and  $y$ , gives 84.31 units increase in

TABLE VIII

Age.....	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Increase.....	100	84.31	76.20	67.48	60.65	54.59	48.36	44.66	39.10	36.03	32.02	29.36	26.55	24.75	20.68	
Total.....	100	184.31	260.51	327.99	388.64	443.23	491.59	536.25	575.35	611.38	643.40	672.76	699.31	724.06	744.75	

growth from 1 to 2 years, and a total of 184.31 units at 2.

To get the increase from 2 to 3, we have  $2 + \frac{.74y - 184.31}{3} = .7037$ , where  $y$  equals the total number of units of growth at 3,  $x$  the increment from 2 to 3, and  $2 + \frac{.74y - 184.31}{x}$  equals the mental age at age 3, which divided by the age 3 equals the I. Q. of .7037 at age 3. Here  $y = 184.31 + x$ . Solving again for  $x$  and  $y$ , gives 76.20 units of increase from 2 to 3, and a total of 260.51 units at age 3. Repeating this process, gives the following median growth curve in terms of absolute units of growth, and which will result in the decline in the I. Q. from .74 at birth to .5585 at the age of 15, as found for the moron grade and given in Table VII.

It is seen that this absolute median growth curve decreases from 100 units of growth a year to 20.68 units at the age of 14 to 15. Had either of the other grades instead of the moron grade been chosen and the absolute median growth curve constructed in like manner the general character of this curve would have remained the same. We may next adopt this growth curve tentatively as correct, and determine the resultant I. Q.'s for different grades of intelligence.

Table IX gives the resulting I. Q.'s at age 10 for different grades of intelligence from 1.20 down to .20 in ten point steps. These are derived with the use of the median growth curve given in the figures of Table VIII. It is assumed again that the ratio of total units of development of any grade to the total median units at the same age remains constant. That is, if it is .50 of the median number of units at age 1, it will be .50 of the median at any other age.

TABLE IX

Grade.....	1.20	1.10	1.00	.90	.80	.70
I. Q. at age 10....	1.45	1.20	1.00	.84	.70	.57
No. points change	+.35	+.10	0	-.06	-.10	-.13

TABLE IX—Continued

Grade.....	.60	.50	.40	.30	.20
I. Q. at age 10.....	.46	.37	.28	.20	.13
No points change.....	-.14	-.13	-.12	-.10	-.07

This shows that the median growth curve that results in the I. Q. changes with age as found with morons will give the I. Q. of other grades, in general, as found by our actual

re-examinations of cases. The I. Q. decreases for the lower grades, and more for the higher grades, up to a certain point beyond which this decrease becomes smaller again, reaching zero, of course, for the median grade of 1.00. It also shows that the I. Q. for grades above the median increases with age instead of decreases. The maximum decrease in Table IX is at grade .60, while as actually found in the re-examinations the I. Q. for borderline cases decreased most. The present data are not adequate for determining at just what grade this maximum decrease in I. Q. occurs with the true median growth curve. The median growth curve constructed to fit the moron grade may be more or less incorrect, because (a) the average yearly decline of 1.21 points for morons may be somewhat incorrect; (b) the decline may vary with age more than is assumed here; (c) the ratio of the subnormal rate of growth to the median rate of growth may not remain constant. It can be shown as a matter of fact that this growth curve is probably considerably wrong, by computing the I. Q.'s at each age from 1 to 15 for the idiot, imbecile, and borderline grades, for this median growth curve, and then compare these I. Q.'s with the I. Q.'s of these grades as actually found.

When these computations are made it is found that the present median growth curve gives I. Q.'s that are increasingly higher, from 1 to 15 years, than the I. Q.'s as found for the borderline grade, and increasingly lower than the I. Q.'s as found for the imbecile grade. For the borderline grade, this difference is  $+.28$  points at the age of fifteen. In like manner, the median growth curve that fits the moron grade as found does not correspond with the median growth curve that fits the I. Q.'s of the borderline grade.

This lack of correspondence must be due to one, two or all three of the factors just noted. If the first two factors are not sufficient to give the amount of discrepancy found here, we are forced to the important conclusion that the ratio between median rate of growth and any subnormal rate does not remain constant. In that case it would become necessary to determine the growth curve for every grade empirically by re-examination at successive ages in order to establish the general tendency for each grade in more detail than the present results do.

The median growth curve constructed so as to give the course of I. Q. changes as found for morons cannot show what changes with increasing age in the rate of decline in the I. Q. may take place for other grades, as it is based on a rate of I. Q. change that remains constant at 1.21 points loss

per year. But, having once established the general type of median growth curve that is required to explain the main facts, we may try out modifications of the curve which are constructed in other ways. The present results indicate fairly definitely at what age mental development ceases. The median growth curve used above shows that the number of units of yearly growth decreases at a somewhat increasing rate with age, but it results in no radical modification of the curve to let it decrease at a constant rate. Assuming now a constant rate of decrease in the number of yearly units of growth, and placing the cessation of mental development at the age of sixteen years, gives a yearly decrease in rate of 6.25 units, when the total number of units of growth from birth to age one is placed again at 100. This results in a total of 844 units at fifteen instead of a total of 744 units for the median growth curve used above. Computing the I. Q.'s at different ages for different grades of cases that result from this median growth curve gives a marked increase with age in the rate of decline of the I. Q. for the higher grades of subnormals. Below grade .40 the rate of decline of the I. Q. becomes practically constant for all ages. This is seen in the following figures, omitting fractions.

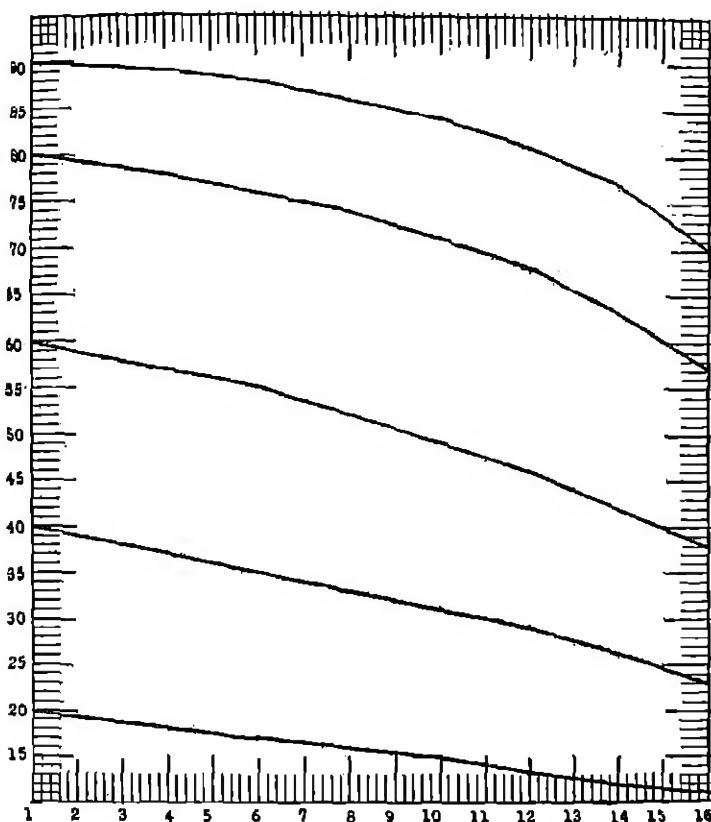
TABLE X

Age	1	4	6	8	10	12	14	16	Av. Yr. Loss
Grade.....	.90	.90	.89	.88	.86	.84	.81	.77	1.31
" .....	.80	.80	.78	.76	.74	.71	.68	.63	1.53
" .....	.60	.60	.57	.55	.52	.49	.46	.42	1.47
" .....	.40	.40	.37	.35	.33	.31	.29	.26	1.13
" .....	.20	.20	.18	.17	.16	.15	.13	.11	.60

The average yearly loss in I. Q. that results from this median growth curve is given in the last column of figures on the right. It is seen that the maximum decline occurs for grade .80, twenty points higher than for the median growth curve that fits the moron grade I. Q.'s. In other words, a relatively small change in the exact nature of the median growth curve shifts this point of maximum decline in the I. Q. over a wide range. The following graphs are for the figures in Table X.

#### Summary

1. The mental ages of the feeble-minded increase with age at a rate proportionate to the degree of mental deficiency.
2. On the whole, the mental age ceases to increase between the ages of fifteen and eighteen, the idiot grade ceasing development about three years earlier than the borderline grade.



3. For the cases figuring in the present study and for the time period they were under observation, 4.8 per cent of the cases gain twelve months or more in mental age in a year, 68 per cent gain from one to eleven months a year, 11 per cent gain or lose less than one month a year, and 16 per cent lose from one to seven months a year in mental age.

4. The lower grades lose more frequently in mental age than the higher grades.

5. The frequency in loss in mental age increases with age, independently of grade.

6. The intelligence quotient decreases with age, and more for the higher than for the lower grades. In the present study the average yearly decrease is 2.19 points for the borderline cases, 1.21 for morons, 1.04 for imbeciles, and .37 for idiots.

7. The rate of decline of the intelligence quotient increases some with age.

8. The several traits of the intelligence quotient are all due chiefly or possibly entirely to a decrease with age in the yearly increment in mental growth of average normal children, as measured in terms of absolute units, and follow mathematically from the nature of this median rate of growth. They may in some measure be due to the scale of tests measuring increasingly too low with increasing age, because this scale is based on the average abilities at each age of school children from which the lower grades have been more and more eliminated by the schools with increasing age.

9. It is probable that even if mental development were measured in terms of absolute units of growth the ratio between the median rate and any other rate above or below it will not remain constant from year to year.

10. The intelligence quotient of cases above the average will increase with age instead of decrease, for the same reason that it will decrease for cases below average.



## PICTORIAL COMPLETION TEST II<sup>1</sup>

By WILLIAM HEALY, Judge Baker Foundation, Boston

### INTRODUCTION

It seems clear that any fair attempt to measure intelligence must include well-directed estimations of apperceptive abilities. Surely the capacity for "putting two-and-two together" in the realm of thought, the ability to "turn things over in the mind," to "rationalize" about perceptual material with what may be drawn on from the ideational and other memory stores of the mind is of the very essence of intelligence and of very great importance in civilization. This, if anything, may be spoken of as a higher mental power.

My interest in a pictorial completion test dates from appreciation of the probable great value of testing apperceptions and from the realization that the language completion test (as used by Ebbinghaus and developed since by others) involves several features which interfere greatly with its early intended use as a test for apperceptive abilities. First, there is the outstanding fact of the immense differences in the opportunities of individuals for acquiring facility with the use of our language, not only differences in formal educational advantages, in family and other cultural opportunities, but also, here in America, in the extent to which English is actually spoken at home or—as unfortunately with some of our immigrant peoples wedded to their own institutions—even in school.

---

<sup>1</sup> This test was exhibited first at the December, 1917, meeting of the American Psychological Association. Soon afterwards the War Department requested the use of it, and, of course, further publication at that time concerning the test became undesirable, even of the scoring values and norms which were at once worked up for the purposes of the Army psychological examinations when the test became part of them. Then the illustration was copyrighted in its several parts in 1918 in attempt to prevent the unfortunate newspaper publicity which befell my first Completion Test. The Judge Baker Foundation (40 Court St., Boston) has recently undertaken to have the test conveniently mounted and boxed and obtainable for a reasonable price.

(Since this article, including the above foot-note, was written, I have observed that this test, including the small plate of the whole series of pictures, which we have purposely refrained from publishing ourselves, and a copy of the scoring values calculated by us, has been published without any mention of authorship in a book entitled, "Army Mental Tests," by Yoakum and Yerkes; Holt & Company.)

On account of this total situation it is impossible to utilize the language completion test as an apperception test equally fair for all.

And whenever the verbal passage becomes really difficult, as in the harder parts of language completion tests, unusual words and those with special shades of meaning increase the difficulties for interpretation of the test results, throwing the advantage strongly to the individual who has had specially good training in language.

Second, I stated my conviction long since, giving instances bearing on the point, that language ability was to be regarded as a specialized ability, sometimes very poorly correlated with the findings on other intelligence tests. We now have considerable data bearing on this. Just as one may possess to a superior or very inferior degree the power of musical expression, so language may "flow trippingly" from the mind or one may by nature halt with words. In the direction of either extreme there is vitiation of the language completion test as indicative of apperceptive ability in general.

Third, the number of details necessary for a desirably complex situation for testing apperceptions can only be presented to a reader at the expense of taking considerable time for the reading and, through this, there is involved another well-known variable, the memory powers or memory span, in this case the memory span for ideas, as they are represented one after another in a long passage.

It suggested itself, then, that a picture completion test might be much fairer as a test for apperceptions, eliminating to a large degree the factors of training and of special ability or disability in language which, as a mode of expression, could be only incidental to the main purpose of the language apperception test. And a well devised pictorial test certainly would give opportunity for presentation of a wealth and exactness of ideational detail at a glance which would require pages of text adequately to cover; as complex a situation as desirable could be depicted without any strain on the memory or any troubled traveling of the attention back and forth, as over the printed page.

In a picture we could present ideas of a simplicity comparable to those of a simple short sentence, or one could illustrate situations too complex for any ordinary rapid solution. And then, as I demonstrate in Pictorial Test II, by the use of the serial pictures one can carry along ideas as they were carried along in the earlier language completion tests where words were elided from a narrative of some length.

Ideas from which to choose in filling the spaces of incompleting sense are represented by a stock of small pictures, and, as one selects verbal ideas from his mental stock in the language completion test, one chooses pictures to fill in what has been cut out of the main illustration.

#### DEVELOPMENT OF THE TEST

The first pictorial completion test which I devised some years ago<sup>2</sup> soon became widely used and commented on, particularly in the monograph devoted to it by Pintner and Anderson.<sup>3</sup> That it seemed to others to embody a sound idea for a good mental test, and that this first picture showed itself to be crude in several features, both for interpretation and scoring, stimulated me to develop plans for a much better test, of the same sort, one that should involve both easy and decidedly difficult problems for apperception. This appeared to be possible by thinking out a pictorial narrative, series of pictures with elements for interpretation carried over from one picture to another just as elements of significance for understanding a narrative are carried over from a sentence or paragraph to later passages, easy or more difficult.

The fact that this idea for the test has not been carried out by others is very likely due to the difficulties and the amount of effort involved.<sup>4</sup> For the development of a series

<sup>2</sup> Described in *The Psychological Review*, May, 1914.

<sup>3</sup> The Picture Completion Test, by Rudolf Pintner and Margaret M. Anderson, pp. 101; Warwick and York, Baltimore; 1917.

<sup>4</sup> All told, a very considerable task has been involved in the development of this second pictorial test, an effort quite beyond the ordinary capacities of any one or two persons. It has been through the fine cooperation of a group, the members of which have given without stint of their time, that the work has been done. Dr. Bronner has been foremost in advice and effort at every stage of the practical development of the test. Indeed, three of the eleven parts of the picture are entirely her suggestions, embodying better ideas than I had been able to advance. The illustrator, Miss Marguerite Davis, brought to the task unflagging zeal as well as artistic skill. The important mathematical suggestions and computations for scoring values were made by Mr. Herbert Sturgis. The entire staff of the Judge Baker Foundation gave itself over to obtaining results on the test both on many trial pictures and on the test as completed, for the establishment of norms. And groups of various nationalities and ages and grades of educational advantages were tested. Much work was undertaken in the interests of the army psychological service during the early part of 1918 in order to establish values for scoring and norms of performance for various groups on this test. Very valuable service in this was rendered by Margaret Fitz and May Bere. A number of outsiders also aided at different stages, a notable amount of work being done by Miss Katherine Coveney.

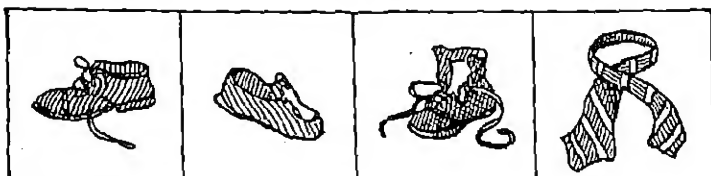
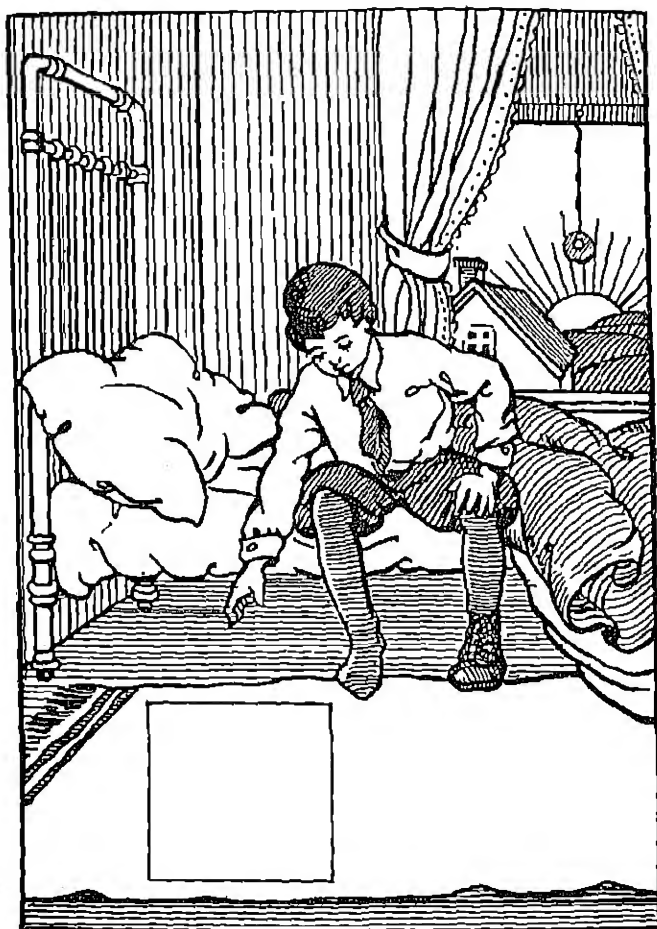


PLATE I

Sketch from the colored illustration of the simplest picture in the series, the demonstration picture, with the inserts which particularly relate to this picture. Of course only one is quite correct, but the others can be inserted with some show of reasonableness.

of satisfactory ideas a considerable task is demanded of the psychologist, preceding combined work with the artist. One of the main practical obstacles is found in the necessity for devising the correct proportionate size and an insignificant background for all the elided parts, so that there shall be no fitting in of the pieces according to the form of the piece, or by joining together parts of the picture. So far as the size of the pictured object for placement and the immediate background is concerned, any one of the related stock of idea-objects must seem reasonably placeable in any one of the spaces to be filled. Then, too, the carrying over of a central idea from picture to picture with broad illustrative conceptions, but with an exactness of representation that shall satisfy critical interpretation is no easy matter.

#### DESCRIPTION

Pictorial Completion Test II consists of an illustration to be divided for convenience into two or three parts when mounted. It is composed of a series of eleven pictures  $5 \times 3\frac{1}{2}$  inches each (one, very simple in idea, merely for demonstration purposes), representing in sequence situations or events occurring during a day in the life of the boy who is depicted in each picture. From each picture there is a piece cut out, one inch square; the picture to be completed by selection of a square from the accompanying sixty small illustrations, only one of which represents the object which is needed to exactly *fulfill the idea of the given picture. In every instance several* of the small pieces depict objects designedly logical in part according to whether various ideas suggested by the given picture are taken into account, but not wholly correct, and hence may partially comply with the requirements for thinking out the situation as pictured. A few of the inserts are purposively suggestive. Of course, most of the sixty small pieces are quite inconsistent with the logical needs of any given one of the eleven pictures.

The ten test pictures present situations planned to be of greatly differing difficulties of solution, and they have worked out so well that children of eight years generally get a credit score, while many adults, even of very superior achievement, find it difficult to achieve a perfect score.

The drawing of the illustrations is satisfactory and the bright coloring stands out well as done by the three-color process. The picture is varnished and should be mounted on three-ply scroll-saw wood. The sixty pieces are to be numbered on the back for scoring purposes and are given standard

placings for presentation, according to numbers to be placed on the floor of the box in which they are kept. The pieces which with some reason might belong in any given picture have a certain amount of grouping in our arrangement so that the search for related "ideas" may not be unduly difficult.

(For those who get the illustration not cut up it should be stated that the numbering of the 60 pieces proceeds from the left top across the page and is repeated for each one of the six lines of 10 pictures. Thus, No. 1 is the scissors, No. 10 the picture and frame, No. 60 the top. Of the larger pictures, the boy getting up in the morning is the demonstration piece; No. 1 is the breakfast scene; No. 2, on the way to school; No. 3, at the school house steps; No. 4, the dressing room; No. 5, the schoolroom; No. 6, boys looking up at the sky; No. 7, the street scene with the policeman; No. 8, returning from the baseball game; No. 9, two boys on the side-walk; No. 10, evening scene in the home.)

#### DIRECTIONS

The entire picture, parts 1 to 5 forming the top row and parts 6 to 10 below them, should be placed in a good light before the subject. The sixty small pieces in the standardized arrangement are presented in the most convenient place where the subject may see them well.

The procedure is then very simple, but should be according to *réglé*. For very dull persons repetition and increased emphasis of the following verbal directions may be needed.

"Here is a picture—it begins here" (pointing to demonstration picture) "where the boy is getting dressed. It shows the same boy, remember, the same boy, doing one thing after another during the same day." (Point along the first row, then along the second, to indicate clearly the sequence in which the pictures come.) "You see in each picture a piece is missing. Here" (pointing to them) "are a lot of small pieces, they fit in any of the spaces. But there are more pieces than you can use. The point is to pick out the piece that you think is needed, that is best, to complete the sense of the picture. For instance, What is gone here?" (pointing to demonstration picture). "Yes, a shoe." If incorrect answer is given, which is very rare, Examiner says, "No, he is dressing and he is stooping for his other shoe." "Now, which is the shoe that he must have?" If correct shoe is selected

Examiner says "Yes. This one" (pointing to low shoe) "wouldn't be right because he must have a high shoe to match the other one."

"Now, that is the way each is to be done. There is always some piece that is the very best one; you can tell which it is by studying the picture. Now go ahead."

*The main point is to give no help after the first explanation.* Nothing further is said while pieces are being inserted except that examinee, if he asks, is told that he may change pieces. After all empty spaces have been filled examiner asks if examinee is quite satisfied, if not, he is told that he may make further changes. When examinee indicates that he has finished as well as he can, time is recorded.

Only the final placings are scored, but record may be kept, if desired, of the changes. Next to the numbers of the pictures in order may be recorded the number or name of the piece inserted.

Twenty minutes is maximum time allowed.<sup>5</sup>

The total score on the test is the sum of the values of the pieces as finally placed. If all are correct the score is 100.

#### SCORING VALUES

The value of minus 5 is to be given to all placings where in the table below no numbers are inserted. These placings represent marked absurdities. The numbers in bold type are for correct placings.

Questions will determine in suspected cases whether a correct placing has been accidental. It happens rarely, as when the correct window in IV is placed, although "any window will do." The score of that placing, obviously, should be no more than the verbal judgment calls for, e.g., the value of an incorrect window in IV.

---

<sup>5</sup> In the army examinations only 10 minutes was allowed on account of necessity for saving time. This was rather unfortunate, although, as a matter of fact, it is decidedly rare that longer than this time is desired by the subject. However, the differences that there may be is shown by the fact that one of the most brilliant men in America took the full 20 minutes and wanted more in which to finish Picture 10, finally leaving a piece in place that gave a total score of 90—the relation of the sewing to the meaning of the picture he had not yet apperceived.

TABLE I

VALUES OF PIECES AS PLACED IN PICTURES I TO X

Pieces	I	II	III	IV	V	VI	VII	VIII	IX	X
1.					2					12.5
2.	0	0	1	2		0	0	0	0	0
3.							1	15.5		
4.					0					6
5.							0	0	0	
6.	0			0						
7.	0	0	1	2		0	0	0	0	0
8.	6			2						
9.		5	0		0					0
10.	0			1						
11.	1			8						
12.		0			0					1
13.		5	0		3					1
14.							1	6		
15.			1			1	0	0	0	
16.							0	0	0	
17.							1	6		
18.			0			9.5		0		
19.		2			0					0
20.										
21.		0			1					0
22.					2					6
23.	1			18						
24.					0					6
25.							2	0	0	
26.			0			4	0	0	0	
27.		5	0		1					0
28.			0			4	0	0	0	
29.										0
30.	2			2						
31.	1			8						
32.		0			7					0
33.							1	7		
34.							5.5	0	0	
35.			11		0	0				0
36.			1			2	0	0	0	
37.		0	4		0	0				0
38.										0
39.							0	0	1	
40.										1
41.		0			2					0
42.	0			2						
43.		0			2					0
44.			1			1	0	0	0	
45.	3			2						
46.	0	0	6	0	0	0				0
47.		5	0		1					2
48.							0	0	5	
49.		10	0		1					2
.....	0						0	0	1	



TABLE I—Continued

Pieces	I	II	III	IV	V	VI	VII	VIII	IX	X
51.....	0	0	1	2	.....	0	0	0	0	0
52.....	1	.....	.....	8	.....	.....	.....	.....	.....	.....
53.....	0	0	1	2	.....	0	0	0	0	0
54.....	.....	.....	.....	.....	.....	.....	0	0	0	.....
55.....	.....	.....	6	.....	0	0	.....	.....	.....	0
56.....	.....	.....	.....	.....	.....	.....	2	0	0	.....
57.....	.....	.....	.....	.....	.....	.....	0	0	1	.....
58.....	0	0	1	2	.....	0	0	0	0	0
59.....	.....	.....	.....	.....	.....	.....	0	0	2	.....
60.....	.....	0	2	.....	0	0	.....	.....	.....	0

The above scoring values were computed from the results obtained by testing 1,542 individuals consecutively, all of them presumably of normal mentality. We did not include records of the feeble-minded or psychotic.<sup>6</sup>

The task of working out proportionate scoring values of correct placements upon the basis of all pieces accurately placed, giving a score of 100, was assigned to Mr. Herbert Sturges. He worked first with a series of 257 records and then with 1,542, including the first 257. The resultant values differed very slightly for the two series, indeed, five of the correct pieces remained practically unchanged, less than .5 each, the greatest alteration being from 12.35 to 11.24 as the value of the correct placing in picture 3.

The method applied to the problem was that of weighting each correct solution inversely to the frequency with which it was correctly solved. The working out of the method is as follows: Opposite each frequency write its reciprocals; find the percentage that each reciprocal is of the sum of the reciprocals.

The rule stated is: Value *a* for problem A is the percentage that the reciprocal of the frequency of correct solution of A is of the sum of the reciprocals of all the frequencies of correct solutions.

<sup>6</sup> Here I wish to acknowledge the value of Pintner and Anderson's criticism concerning the inadequacy of the scoring methods and norms for my first completion test. In turn, it was this criticism that challenged Dr. Bronner and me to attempt to develop a less clumsy system of scoring than the one devised by these authors. The main point they make is thoroughly valid; many tests do need standardizing for age and other groups. But on all except group tests the necessary labor—over a thousand hours were devoted to standardizing this our present test—and the requirement for much coöperation stands in the way.

TABLE II

VALUES FOR A SERIES OF 1542 RECORDS

Picture	Correct Placings	Percentage of total correct placings	Reciprocal of frequency	Score	P. E. of Score
1	1130	14.44	0.000885	6. (5.77)	0.06
2	655	8.37	0.001527	10. (9.96)	0.20
3	580	7.41	0.001724	11. (11.24)	0.25
4	363	4.64	0.002755	18. (17.96)	0.56
5	925	11.82	0.001086	7. (7.08)	0.10
6	683	8.73	0.001464	9.5 (9.55)	0.18
7	1228	15.69	0.000814	5.5 (5.31)	0.05
8	416	5.32	0.002404	15.5 (15.67)	0.46
9	1325	16.93	0.000755	5. (4.92)	0.03
10	520	6.65	0.001923	12.5 (12.54)	0.30
Totals	7825	100.00	0.015337	100.00	

The question of giving values to solutions other than the one correct solution of each picture could not be answered by mathematical considerations based on the frequency of such solutions. Arbitrary determinations embodying as much common sense as possible were invoked. No approach to correctness, it was felt, could be worth more than half as much as the correct solution. So the pieces which were planned to represent ideas similar to the single correct piece but missing the real point were given values of half or a little less than half. Thus the value for correct placing, the red book, in Picture 2 is 10 (the nearest half decimal to 9.96). But if one finally chooses the blue book or the two books or the lunch box or the box of pencils the score is 5. Of course it could be nothing else than the red book if one apperceives the situation, the boy having dropped something out of his strap, and his books, plainly colored, being shown strapped together in the previous picture. The only other possibility is a foot-print on the ground, and that receives a score of 2. In the other pictures the commonsense of the situation and the possibilities differ and various arbitrary rulings concerning scores have been made accordingly—always based, however, on the original computations of correct scoring values.

If a piece of indifferent meaning, such as a blank in Picture 2, is left in place, or if, as extremely rarely, the space is left unfilled it has been decided that the score should be 0. But a large share of the possible placings are obvious absurdities and logically should be penalized. It is pretty clear just which these absurdities are. It was decided that minus 5 should be the standard score for these absurdities.

Though the scoring of these latter classes of placings is thus numerically arbitrary, still it must be remembered that the scoring as a whole is based upon mathematical computations of fair values according to ascertained relative differences in performance and that once this was done, the minor values as established are as fair for one person as another. (Of course I realize the suggestion value or the association value of any piece is never the same for one person as for another, but the same might be said to a greater or lesser degree about almost any mental test.)

TABLE III

## PERCENTILES FROM THE ABOVE CHART

Ages	7	8	9	10	11	12	13	14	15	16	17-20	20-50
75 percentile	24	41	48	59	63	66	69	72	76	75	76	78
Median	9	27	37	47	54	55	58	62	64	66	65	65
25 percentile	-6	7	23	32	41	45	50	52	54	54	54	54

*Special educational advantages.* Whatever the findings for those who have had educational advantages above high school it would hardly be safe to allege that any good score would be due to such advantages rather than to native ability. It is more than conceivable that a considerable number of those who go further with their schooling do so because of their already demonstrated ability. Nevertheless it was felt to be much wiser to keep separately the 231 records of this group.

TABLE IV

Scores Obtained	No. of Subjects
91-100	57
81-90	54
71-80	47
61-70	42
51-60	21
41-50	8
31-40	2
	<hr/>
	231

75 percentile = score of 90

Median = score of 80

25 percentile = score of 68

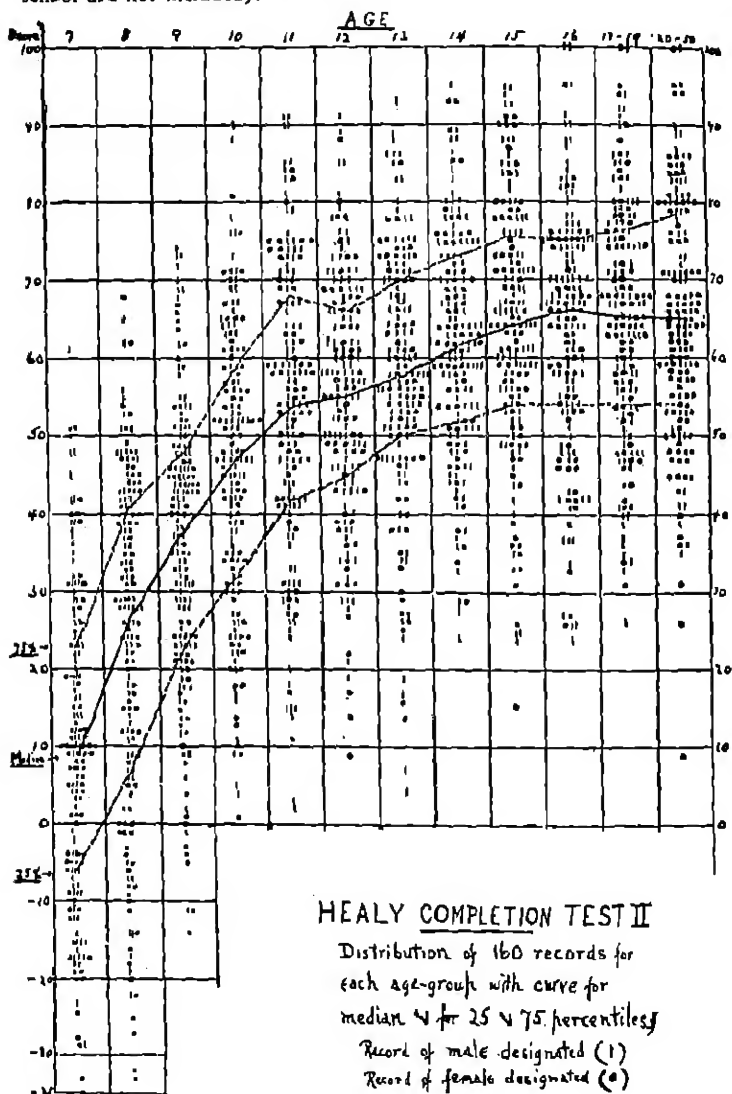
It is interesting to compare these much higher percentiles with those of the main group. And then it should be stated that a perfect score was obtained by 25 (11 per cent) in this last series as against only 4 (1.2 per cent) of 320 of the same age-group (above 17 years) in the main groups.

## NORMS OF PERFORMANCE

Age-groups

FIGURE 1

Age distribution of total scores, 160 for each age-group (records of the feeble-minded and of those with educational advantages above high school are not included).



While complete significance cannot, of course, be given to the findings for age-periods with the higher educated group on account of uneven and too few numbers, yet it is striking that 20 per cent of 65 individuals 35-50 years of age gave perfect scores and that their median score was 85, while 7.3 per cent of 164 with ages 19-35 did perfectly, with a median score of 76. It would seem as if among well educated people youth did not have the best of it in this test.

*The Mentally Defective.* Records of the performance of the feeble-minded on any one test show the futility of attempting to establish for any single test norms which are characteristic of the mentally defective. Mental defectives differ greatly among themselves in their special capacities and in other important features of mental life than are expressed by general age-level ratings on a given "intelligence scale." Some are feeble-minded because of certain defective memory powers, others with good memories have poor reasoning ability, some show great unevenness in their mental powers, some are below par evenly in their various mental capacities. When really good chapters on analytic psychology of mental defectives shall have been written, we may learn something of the complexity that most certainly exists in the mental life of the feeble-minded, which to many seems a simple affair. As it stands, we should scan the records of defectives in general on any special test with the knowledge that they represent the performances of individuals with age-level ratings differing widely from actual age experiences and from varying world experiences. Or the records may come from those who are not clearly feeble-minded except insofar as social behavior is concerned (institutions for defectives sometimes retain such cases) or from schizophrenics who are not infrequently held among the feeble-minded.

Even on a test involving such complex mental functioning as the one under consideration we may find great differences among defectives, at least of the higher grade, who have equal rank on age-level scales. However, the ordinary observation that some show clear mental grasp (apperception) of their own disabilities and various other practical situations, and other defectives of equal age-level rating by the Binet scale or its modifications fail thus to size up the real situation in regard to themselves and other affairs, ought to make the fact of their mental difference in this particular altogether plain and suggest at the outset that their scores on a special test for apperception might differ widely.

*Other Groups.* From records obtained by way of experiment, but too few for safe deductions, we have noted several suggestions for special studies. Persons who have seen very little of books and are *unaccustomed to observe pictures* understandingly or very little in any way, all of them being, with us, illiterate foreigners, were at a great disadvantage and clearly could not interpret the sky or the interior of a room as such, to say nothing of finer points in the pictures. *Psychotic* individuals have made all sorts of records, as may be supposed—any study of their performance would have to be by types and stages of psychoses. The apperceptive ability of *old age* is a matter of special interest and perhaps might turn out to be of great practical importance, because of the value that is given by tradition and social usage to the opinion of men of age, especially as they function on the bench or in administrative positions.

#### DISCUSSION

The universal interest evoked by the test is readily understandable. As a series of incomplete pictures it challenges the performance of everyone, for even the mentally defective, above low grades, see enough of the significance of what is depicted so that they become interested and are able to score as having observed correctly and reasoned correctly about some of the material in the pictures. An opportunity like that of everyday life is afforded them for some sort of not altogether unreasonable performance. In this, namely, in developing a test that in given situations for apperception should have simpler elements capable of being readily rationalized by very immature minds and at the same time embodying features of much greater difficulty, up to those relationships which require close thought by competent reasoners, one of the foremost aims of this test has been realized. And it is a fact that even the cleverest, when they note the possibilities of interpretation, enjoy working out the problems of the test as tasks calling for good effort.<sup>7</sup>

The ideas as depicted in the pictures finally drawn are so clear that they present no logically equivocal issues. This has been proved for us not only by the records of performance, but also by the comments of those doing the tests and their acknowledgments as they have later seen the point for correct solution; even after having themselves earlier failed to apper-

<sup>7</sup> We have just heard that the work with French children on this test as done by Marion Pilpel through Simon's clinic in Paris has been very satisfactory. The situations have been readily grasped by these children.

ceive some essentials. Indeed, it is interesting to note that in the two pictures awarded the highest values because of the infrequency of their correct solutions, the correct insert is indicated with more exactitude by objects clearly shown in the main pictures than is the case in any of the other pictures. There is less chance of equivocation in these two pictures because the possibilities of solution, once grasped, are so immediately related to the factual necessities of the situation. The most common failures, then, are directly perceptual and inferential in character, involving faulty observation or faulty reasoning concerning objects and phenomena universally known and among the most plainly observable in the illustrations.

In even the simplest analysis of the test, as a test measuring something much more than primary mental powers, such as memory, we encounter the facts, then, of its easily observable and readily recognizable material, of the facile perceptibility of the elements in and for themselves. The difficulty of the problem comes in rationalizing the value of each whole situation, of the worth of the various elements and of the cut-out objects (the store of ideas) as possible solutions in the given situation—and this is *apperception*.

Through its presentation of problems that are visual, non-language, ideational, pictorial, a representation of everyday scenes and activities, this test brings into play an important mental ability—whether specialized as being related to observational and pictorial material or capable of wider interpretation—an ability that we would eagerly know more about for many practical purposes. The reaction expressed by one well-known psychologist on first seeing the test, "Well, if I were hiring people to assist me, I think I should like this test to hire them by," may be paralleled by many another opinion of like import from those who have been using this among other tests. Noting that the higher records on the test, we may take the perfect score, 100, for example, have been done by only very occasional youngsters of high ability, as proved by other tests, and by some adults who have established reputations for achievement, and these records have been equalled very rarely indeed by others, for instance, by a domestic cook, we may look further and generally discover in these cases (as in the latter instance where this cook was found to be well known as having exceptionally good common sense and judgment) that there are obtainable evidences of the correlation of such a performance with otherwise known good mental capacities. The powers that this test in some measure gauges are certainly those allied with better human achievement.

## THE CONSTRUCTION OF TESTS FOR THE DISCOVERY OF VOCATIONAL FITNESS

By FRANK WATTS, *University of Manchester*

It will be agreed that the successful construction of tests for the discovery of vocational fitness must depend in the long run upon an accurate psychological analysis of the various occupations followed by men and women. Only when this has been accomplished in a sufficient number of cases will it be possible to formulate general principles in accordance with which we may know how to proceed in given instances to frame reliable tests for ready use. The attempt to elucidate such general principles may at the moment appear to many to be premature, but it is difficult to spend much time in the work of trying out tests in industry without feeling that the fundamental assumptions underlying the different methods of test-construction call for careful examination. We cannot hope to proceed far in practice with safety without guiding principles; and it is with certain of the theoretical aspects of this subject, therefore, that I propose to deal in this paper. Occasional reviews of the work already done cannot but be helpful to those engaged in actual investigations.

How may we classify the tests already in use? One broad method of classifying them would involve their division into analytic and synthetic tests, the former representing the method of investigating separately the various mental and physical factors which enter into the performance of a piece of work, while the latter type stands for the attempt to measure vocational aptitude as a unitary capacity. Synthetic tests may be further subdivided into the various sample tests such as (1) vocational miniature tests, oral trade tests, picture trade tests, written trade tests and performance trade tests; (2) empiric tests, which at present represent the application of the method of trial-and-error in vocational psychological psychology; and (3) analogous tests which, combining the advantages of the analytic and the sample tests, will, I venture to believe, prove the only satisfactory means in the long run for diagnosing potential as distinct from actual vocational aptitude of a non-routine kind. Let me describe what I conceive to be the advantages and the disadvantages of each



type of test as far as experience permits me safely to do so. I shall venture to take a point of view which is slightly different from that of other workers.<sup>1</sup>

The analytic test, as has been indicated already, attempts to isolate and measure separately in respect of individual applicants for employment those physical and mental factors which are predominantly demanded in the given occupation. There are obvious limitations to the utility of this method. One could hardly employ it to take an extreme example from outside our subject, in determining whether a man (or woman) were suitable for the life-occupation of husband (or wife). Firstly, it would be well-nigh impossible to analyse out into clearly distinct factors all the various mental capacities called for in the state of matrimony. Secondly, if such an analysis were achieved it would be practically useless to attempt to isolate each factor involved for separate investigation and measurement; for how could one be sure of dealing with any one quality with the certainty that it was completely detached from others intimately associated with it? Thirdly, even if these two insuperable difficulties were removed we should still need to know in what proportions the single elements entered into the composition of the complete capacity, and this raises an important problem which I have never seen faced by anyone who has recommended or employed the analytic method of testing vocational fitness. This problem is one of deciding how to "weight" the scores for the various single capacities so that the total score is really a representative score. I have already indicated in another paper that by "weighting" the scores for important capacities in a number of subjects, rankings may be obtained quite different from those obtained when weighting is neglected.<sup>2</sup> Fourthly, to continue our objections, the analytic method assumes that a psychic whole is merely the sum of the factors into which it can be conveniently split up, and this assumption ought not to be accepted in the absence of proof. The belief which would seem to be justified is that personality is not thus arbitrarily to be divided up into part capacities without loss. To

<sup>1</sup> Cf. Essay by Cyril Burt in *Lectures on Industrial Administration*, pp. 79-104. H. Hollingworth, *Vocational Psychology*, ch. v. B. Muscio, *Lectures in Industrial Psychology*, pp. 119-152.

<sup>2</sup> See *The Outlook for Vocational Psychology*, *Brit. Jour. Psych.*, XI, 2. A simple method would be to allot more marks to some of the tests in any series than to others. As a possible aid to an elucidation of the problem which would then arise of the relative importance of the various capacities entering into a performance a hierarchy of our human capacities is appended.

use a homely illustration, an unanalysaltic factor of experience may result in two women producing two entirely different cakes with similar ingredients mixed in the same proportions, the baking conditions being constant.

Professor Seashore's patient analysis of the requirements of the singer into something like forty different qualities, though an ingenious piece of work from one point of view, would hardly seem, therefore, to be useful to the vocational psychologist in search of a test of singing ability. Separate investigations by scientific methods regarding the range of voice, the ability to sustain tones, the memory span, the delicacy of pitch discrimination, the association type, "mental plasticity," creative imagination, and so on, would prove to be unnecessarily tedious and probably highly unsatisfactory in view of the difficulties already indicated.

Let us take two potential singers A and B who may have happened each to score a hundred marks in a series of tests applied in accordance with the analytic method. Unless they have made identical scores in the separate tests—which is unlikely—does it help us to say they have been "scientifically" proved to possess equal ability or promise? A has scored heavily in one direction because he possesses a magnificent natural voice, but B who lacks such an asset has made up for this deficiency by the possession of marked creative imagination and control of feeling. So that all we can gather from the elaborate tables of results would be, what we probably knew already, that A excelled in one way and B in another.

There seems to be justification, then, for suspecting that in all but the simplest cases the analytic method will probably prove worthless. It is where routine work is concerned and the chief capacities called for are no more than one or two of the sensory or motor functions which are almost mechanical and practically divorced from the higher forms of intelligence, that the analytic test may be useful, and useful here because the intellectual factor does not complicate significantly the various capacities measured. But even then there may be dangers in ignoring in any series of tests the intellectual factor as we shall hope to show later.

The method of "the empiric test" is strongly recommended by Hollingworth (who, however, admits it to be a "rough provisional and unanalysed expedient") as representing "the most promising experimental procedure for the immediate present and perhaps for some time to come."<sup>3</sup>

---

<sup>3</sup> P. 119, *op. cit.*

It is a method which necessitates no preliminary psychological analysis of tasks. The experimenter simply applies a number of miscellaneous standardised tests, e. g., the *completion test*, the *analogies test*, the *cancellation test*, etc., and notes how closely the scores obtained correspond with estimates of trade ability. Those tests which correlate highly with such trade estimates are then accepted as suitable for diagnosing capacity for the occupation under consideration.

"Perhaps the most perfect example of this purely empirical procedure," writes Hollingworth, "is the investigation which has now been conducted for several years by Mrs. Woolley and her co-workers in Cincinnati. Children who leave the grades to enter directly into some sort of industrial occupation, are examined by a miscellaneous assortment of simple mental tests. These records are preserved and the subsequent successes or failures of the pupils in the various sorts of work undertaken by them in later life are as carefully recorded as possible. It is hoped that when a sufficient amount of material of this nature has been accumulated, the two sets of data may be compared and information thereby secured concerning the relation between ability in the tests and the types and degrees of industrial fitness."<sup>4</sup>

To argue anything of a general nature from this special application of the empiric method is to assume that the interests of boys and girls are either permanent or change in an orderly and predictable manner. But tests which interest the boy or girl while at school and tend therefore to call out a maximum of ability in their performance may cease to attract in later years; so that *if applied later* they must inevitably produce results which are not to be relied upon to the same extent as data for diagnosing vocational aptitude.

But speaking generally, the empiric method is unsatisfactory because it is at best nothing more than a makeshift method for avoiding the difficulties of psychological trade analysis. To continue to employ this method without at the same time making an attempt to psychograph the work concerned is a procedure which will not be commended by any thinking person. The empiric method, more than any other, places us completely at the mercy of *sampling errors* where our human material is concerned, and of the errors, too, of the trade expert upon whom we must rely for trade rankings of the individuals tested. One of the greatest troubles in vocational testing is to find foremen who can give unbiased judgments about the bare vocational aptitude of the workers under them! There is

---

<sup>4</sup> P. 114, *op. cit.*

always the double tendency present, which few but the most highly trained observers can control, to pass a favorable judgment upon those whose temperaments "fit in" readily with their own so as to make for a harmonious working relationship, and also to see specific aptitude in what Thorndike calls a "halo" of general ability, but never apart from it in its own shape and outline.<sup>5</sup> Thus in the vocational world the empiric method of test construction in effect, shows us the blind leading the blind. Only when we are still completely unsure of our vocational analysis should we be content to employ the empiric method of test construction, and even then not blindly but for the deliberate purpose of opening up a way to a further knowledge of the psychological requirements of an occupation.

The various types of sample tests which have been advocated indicate an attempt to move away from the confusion of theory created by the analysts and the empiricists. But we shall see that if the theorists have unduly neglected the practical requirements of industrial tasks the practical men who have devised trade tests have equally neglected to consider the theoretical implications of their procedure. To take first the commonest form of the sample test which has been used, viz.: the trade performance test. Here we see the applicant for employment given an actual sample of the work which he may later be called upon to perform. His fitness or unfitness is to be gauged from an inspection of the results. This is a method which lends itself readily for use in a large number of occupations; for example, in the choice of typists, clerks, musicians, athletes, skilled artisans, etc. The principal conditions for its successful application would seem to be: (1) that, at least, a representative part of the work is suitable for performance as a test without serious modification and (2) that the results can be expressed quantitatively.<sup>6</sup>

The performance test-method is not without its pitfalls. At first sight it may not seem to call for any great psychological insight and yet to proceed without regard for the influences and effects due to unfamiliarity with the test conditions, lack

<sup>5</sup> In framing intelligence tests Binet solved this difficulty by constructing his *échelle métrique* which provided him with an independent standard for individual measurements. The same principle must be applied to industry, and the writer is at present attempting to do this in the case of one trade.

<sup>6</sup> In Dr. J. C. Chapman's book on "Trade Tests," (Holt & Co., New York, 1921), will be found standardized performance tests for carpenters, pipe fitters, sheet metal workers, electricians, machinists and mechanics, blacksmiths, stenographers and typists, chauffeurs and so on.

of practice, fatigue, the use of unstandardised samples and unstandardised instructions, nervousness, etc., all of which can only be negatived by the employment of scientific methods of procedure is to risk serious error. Moreover, we must know before we can judge a sample piece of work what an average performance is, and this will necessitate considerable research in every case. To be successful, therefore, the sample performance test must be applied under controlled conditions and the results interpreted in relation to known standards. An intimate acquaintance with the methods of experimental psychology will alone ensure that these conditions and standards have been obtained.

The vocational miniature test is a special form of the sample test possessing its general disadvantages with particular ones of its own in addition. It calls upon the subjects of experiment to deal with small "toy" representations of the actual work to be done. Thus a switchboard, much reduced in size, may be used for testing telephonists, or signalmen may be required to manipulate tiny levers in response to demands for signals. It must be at once apparent that the number of tasks which can be reproduced in miniature is extremely small. But the method has been adversely criticised on account of its own intrinsic defects. Münsterberg quite rightly says that "a reduced copy of an external apparatus may arouse ideas, feelings, and volitions, which have little in common with the processes of actual life." There is no defence which can seriously be put up against this criticism.

Moreover, as far as physical labor is concerned there is reason to believe that ability to perform the fine movements called for in working with small models is not at all indicative of ability to perform the larger movements of the actual work which the small model is intended to represent. In the two cases not only will different muscular and nervous co-ordinations be necessary, but also different types of interest. Thus the watchmaker and the miniature-painter would usually be completely unsuited "temperamentally," as we say, as well as physically for employment respectively upon steam-turbines and motor-generators, or upon big poster work. In other words, there is probably no "general factor" in motor ability.

The oral trade test, the picture trade test in which a picture of a tool or piece of work is shown, and the written trade test are all types of sample test based on the questionnaire-method. Such tests obviously, are meant to elicit, and can elicit, nothing more than information. If these tests are used

---

<sup>1</sup> *Psychology and Industrial Efficiency*, p. 67.

as supplementary to performance tests they may be employed to serve a useful purpose. But most of us harbour the prejudice that to know all about a game of skill or a trade is by no means proof that one is an expert in it, so that if such tests are used as substitutes for performance tests then the general question is raised as to the relation which exists between skill and knowledge, and whether the latter can be accepted as adequately representative of the former. For the elucidation of this problem everything will depend on the type of question asked and the knowledge presupposed in the individual circumstances. In many cases it is, I believe, a mistake to attempt to measure a bare ability independent of experience, because such experience may be the only proof of the presence of the interest upon which permanent efficiency depends. But it would seem on the whole that this type of test is one which can easily be coached for, if we are to take the American Army oral, picture, and written trade tests as truly typical of the method.

The chief objection, however, which the vocational psychologist will be inclined to raise against all sample tests, of no matter what kind, is that they may measure actual rather than potential capacity. They will not help us to select from among individuals without experience of a task those who would benefit most by training. The type of psychological test which is most urgently demanded—the true vocational test, that is, is the test which can be applied to those about to enter an industry with the idea of gauging their probable fitness in advance, so allowing us to advise the misfits to choose another occupation before it is too late. The only satisfactory test which promises well from this point of view (as a means of choosing workers above the routine level) is the skilfully constructed analogous test. The fact that particular analogous tests have not been altogether successful in the past is no argument against their real value.

The analogous test method calls both for a tentative psychological analysis of the work to be done and for the construction of problems calling out as a combination the essential capacities and interests concerned, in much the same proportion as they are demanded in the actual tasks, but in such a manner as to allow potential capacity, when necessary, to compete on equal terms with capacity already fully developed. Really, therefore, it may often become in practice the sample method purged of its peculiar defects. It will be agreed that the work of constructing such tests is no simple matter, but it will prove, I believe, the only real way of progress.

Münsterberg's famous test for the discovery of car-driving aptitude was, up to a certain point, a test fulfilling these requirements. It was defective, not because of any inherent unsoundness in the analogous method, but because it issued from an incomplete analysis of car-driving aptitude. The method was employed by Münsterberg to replace the unsatisfactory method of vocational miniature which had proved defective. "Not the external similarity of the apparatus," wrote Münsterberg, "but exclusively the inner similarity of the mental attitude,"<sup>a</sup> should be the chief characteristic of a test which is intended to reproduce the conditions of work in any occupation. This inner mental attitude is usually so difficult to analyse that it is not surprising that few excellent vocational tests have so far been constructed. Vocational psychography, that is to say, is still little more than a mere phrase.

What has been attempted so far, however, will enable us, at least, to study our present position. One may say, to sum up the possibilities of the various types of test, that the various sample tests are useful for choosing skilled workers but useless for selecting youths for apprenticeship, that empiric tests should only be used when other methods fail, that analytic tests may be used (subject to reservations) where routine work is concerned, but that where intelligence as well as specific aptitude is necessary the analogous test is needed.

Most of the mistakes and shortcomings characteristic of our early attempts at vocational psychography may now be classified under the headings of errors of inclusion, errors of exclusion, and those errors of proportion already spoken of in connection with analytic tests as responsible partly for their defects.

To consider the errors of inclusion first. Occupations may be grouped into two classes according as to whether intelligence is or is not the predominant factor demanded in them. In a large number of routine occupations the possession of more than an average amount of intelligence makes for dissatisfaction, restlessness, and inefficiency. Yet, especially where the empiric method has been employed, intelligence has been a factor which unwittingly has been allowed to creep into the test performances designed to reveal aptitude for various types of routine work. Tests which are virtually intelligence tests have been applied in various occupations where the workers have happened to be of low grade mental type, so that the sampling factor has obscured the significance of the results. But since differences of intelligence, even if small,

---

<sup>a</sup> P. 68, *op. cit.*

still exist and so are responsible for the test rankings, a close correspondence between test results and trade ability has been found. But this is not to say that universally wherever we have greater intelligence there we shall have also greater permanent aptitude for the routine work under consideration. For let us suppose that the Knox Cube Test, which has been used successfully to discover capacity among girls employed in label pasting, were to come to the notice of a tyrant with wide powers, ruling a modern community, who had no use for teachers of psychology. Would he be right in acting on the assumption that if these persons endowed with marked introspective habits of mind were given the Knox Cube Test to perform then the most successful would make, for more than a few hours, the best label pasters? Obviously not.

Thus with every routine operation, tests will often work well only supposing that (1) they are tests of aptitudes which are largely independent of intellectual ability, and (2) that the sample from whom we are to make our choice contains no individuals of more than average intelligence and initiative. These conditions were given when Mr. S. E. Thompson<sup>a</sup> carried out his reaction-time experiments with 120 girls employed by him as bicycle-ball inspectors so that he was able to depend entirely on speedy reaction-time as the most essential element making for success in their work. Had the work involved a fair measure of intelligence, mere reaction-times would have been much less significant of capacity. As it was he was able to select 35 girls superior to the others in the essential quality, and by a modification of the conditions of work (better wages, shorter work periods, holidays, etc.), was able to produce an output not less than before.

To put the matter in another way. In many occupations it may be discovered that a specific aptitude or native interest plays a greater part in making for success than intelligence. In such a case we must beware of relying entirely on tests which are really tests of intelligence rather than of specific capacity.

The errors of exclusion are usually due to the neglect of a definitely specific interest in a test-appeal. Thus, Münsterberg's test for car-drivers was excellent in conception but intelligence was allowed to play too considerable a part, while the interest in driving real cars in real streets which alone will ensure permanent satisfaction and efficiency in the work did not enter into the test to any extent. Had the test been one in which only those with the interests already referred to, or

---

<sup>a</sup> See F. W. Taylor: *Scientific Management*,



at least the capacity for developing them in the actual work, could have done well the College students who performed the test and excelled above the best drivers would have found their proper level.<sup>10</sup>

The foregoing considerations must suggest to the reader how imperative it is that we should attempt to analyse out as completely as possible the various mental and physical factors which enter, or should enter, into the performance of the principal industrial tasks. Now the uncritical acceptance of text book classifications of the mental factors which may enter into the composition of vocational capacity will not help us very much.

In reading works on the subject of vocational selection, one cannot fail to be struck by the fact that writers are apt to use many of the psychological terms for our various aptitudes in such a way as to blind one to their frequent complexity. Thus, among supposed mental "elements," we find such things as tact, perseverance, attentiveness, reliability, caution, accuracy, mental *tempo*, memory, industry, neatness, etc., and the impression conveyed is that there are definite unitary mental factors easily discoverable corresponding to these vague general terms. The result of this use of general terms, however, is frequently unfortunate. For example, when the peasant ploughs a clean straight furrow, or the seamstress shows a beautifully worked hem, or the typist displays one's correspondence with marked legibility, or the housewife maintains a home in spick and span condition, we are apt to say and think that they all manifest in their own way a single indivisible quality called neatness. Such a way of thinking often deceives us, for these varied forms of neatness though conveniently summed up in this manner under a single name are by no means the same qualities psychologically. Only a crank would expect to discover the best ploughman or the best seamstress by means of neatness in a written test. Similarly, if it were agreed that a proof-reader needed to be gifted for detecting errors it would be absurd to test him by any kind of exercise except one dealing with printed matter.

So when one comes to investigate the mental make-up of adult workers as it is expressed in any particular occupation, what is found is that the "elementary" qualities popularly ascribed to such workers are by no means as simple and indivisible as is implied. We do not find as a result of actual experiments that the best proof-readers are always characterised by the possession of a power of close attention applicable

---

<sup>10</sup> P. 80. *op. cit.*

in many directions, that fitters or instrument-makers display a quick recognition of the significance of form (of no matter what kind), that telephonists develop a good all-round memory, or that street-car drivers exhibit a general ability to apprehend the relations, no matter what, existing between several things, no matter where. What we do find instead is that with the vast majority of adolescents and adults each ability is more likely to be quite clearly specific and limited in scope. That is to say, a person may be a bad observer as regards the common incidents of the streets, but a good observer of natural phenomena; a youth may have a bad memory for dates in history but a good memory for football fixtures; a man may have a remarkably effective power of paying close attention to the movements of machine parts but not to checking figures or letters which pass before him; one may be able to apprehend the relations existing between the many factors involved in a game if one is a spectator, but not if one is a player; a woman may exhibit a remarkably quick mental *tempo* in clerical work, but a deplorably slow *tempo* in organising a home; and so on. Thus it will be clear that no ability can be wholly understood or profitably investigated apart from the living interest which energises it.

The difficulty which will suggest itself to many who have thought about the part which any interest should play in a psychological test is that since most interests do not develop except through the assimilation of knowledge and skill, it is impossible to allow them to enter into the performance of a test without thereby placing their possessors at a great advantage over others equally capable of doing good work whose interests are not yet aroused. It may help to remove this difficulty if we distinguish between the capacity or skill which is determined largely by the existence of an interest or interests which any intelligent person may develop irrespective of his age, and that capacity or skill which is largely due to the development of a specific interest at an early age for the absence of which no amount of intelligence in later life will adequately compensate. Thus if one should be selecting men for responsible open-air work in a distant colony it would probably be much more satisfactory to choose those whose interest in the colonies and in open-air life is of that deep-rooted type which appears early, becomes nourished and strengthened by tales of adventure and danger, and remains proof against the passage of the years. A question such as, "When did you first get to know about this kind of work?" will usually reveal the age and strength of the interest concerned. In many cases

HIERARCHY OF THE HUMAN ABILITIES <sup>11</sup>			
	Processes of the Mechanical level	Processes of the Organic level	Processes of the Psychic Human level (Man as creature)
SENSE	The sense of reactions (measurable in speed and acuity): Taste, Touch, Hearing, Sight, Temperature, Kinæsthesia.	SENSORY DISCRIMINATION or sensitiveness to differences in quantity and quality of sense impressions: e.g. heavy and light, smooth and rough, dark and bright, etc.	The finer discriminations: colour, rhythm, pitch, etc. Ideal sense; of 1. <i>order</i> 2. <i>form</i> 3. <i>beauty</i>
+ INTELLIGENCE		Intuition: <i>Recognition</i> of the practical significance of the above difference.	Logical understanding. <i>Imagery</i> : e.g. visual, auditory etc.
	Habit Memory.	ASSOCIATIVE MEMORY: for practical experiences. Memory as RECOGNITION.	RECOLLECTION: a. Arbitrary Association. b. Controlled Association.
+ FEELING	PLEASURE-PAIN.	The Primary Emotions	The Concrete Sentiments with the emotions: joy, pity, sorrow. (When not rational = <i>prejudices</i> and <i>complexes</i> .)
= WILL	IMPULSE: (Appetition and aversion) <i>HABIT</i> as shown in ROUTINE SKILL Attention, as naturally fixed or fluctuating.	The Primary Instincts (=Native Interests)  ORGANIC SKILL Attention, as determined by native interest.	The fully enlightened WILL of the 1. saint 2. philosopher 3. artist ARTISTIC SKILL

<sup>11</sup> See also *Brit. Jour. Psych.*, Vol. XI, Pt. 1, p.

where industrial efficiency depends upon interests developed early in life it will be found possible to construct "age scales" in the Binet manner so that the extent of the progress made in the developing capacity and skill may together be estimated. But even where occupations demand the functioning of interests which normally do not develop early it will still be necessary to appeal to them in our tests since intelligence tests alone will not suffice. We must still contrive to construct tests which will allow the same interest or interests to operate in the test as in the actual work, and the consequence will then be that those individuals possessing such an interest or interests but lacking "all-round" intellectual ability, can compete on equal terms with those intelligent persons who are able to make good at most occupations interesting to the average man or woman.

A COMPARISON OF BRAHMAN AND PANCHAMA  
CHILDREN IN SOUTH INDIA WITH EACH  
OTHER AND WITH AMERICAN CHILDREN  
BY MEANS OF THE GODDARD FORM  
BOARD

---

By D. S. HERRICK

---

Do racial differences extend to mental characteristics? Are differences in color and other physical features accompanied by differences in general intelligence, or in psycho-motor ability? These are questions of considerable interest that as regards the people of South India have not been approached from a scientific point of view, nor investigated by scientific methods. Anthropometric observations were made some years ago. The material thus obtained was used in the preparation of a work of several volumes, "Castes and Tribes of South India," by Dr. Thurston. Mental differences were not investigated.

During the latter part of 1919 the writer took up the subject. A series of tests were made with the Goddard Form Board in some 20 or more schools in the Madras Presidency. More than 700 children of ages from four to fourteen were examined and the results tabulated. As it seemed to be worth while to compare not only Indian children with American children of corresponding ages, but also different racial groups of Indian children with each other, half of the tests were made on Brahman children and half on Panchama children. The term "Panchama," it should be explained, signifies in South India the very large social group in which are found the very lowest castes. Hindus do not regard them as castes at all in the strict sense of the word, but rather as "outcastes." It was assumed that, so far as there is any distinct racial cleavage among the native peoples of South India, it would be between the Brahmins and the lowest castes, or the outcastes. A fundamental claim of the Brahmins is that their blood is unmixed with that of other races. While it seems probable that in the early centuries of the Aryan migrations there was considerable intermarriage between the invaders and members of the Dravidian race or races that they found occupying the land, it is fairly certain that for many centuries there has

been no general intermingling of blood. Doubtless the Dravidian tribes were of different degrees of civilization. The most backward of them, the jungle dwellers, of whom there are living examples in many parts of India today, are the progenitors of the outcastes of the present day. So far as intermarriage occurred at all, it must have been with the more civilized of the Dravidians.

That there are marked physical differences between most of the Brahmans of South India and the typical outcaste living in the same village is evident from a casual comparison of their color and features. It may also be shown by a comparison of their cranial and nasal indices. Investigations along this line are to be found in the work by Thurston referred to above. Whether there are also marked differences in mentality that can not be accounted for by differences of environment in its broadest aspect is a question of much interest.

The choice of the kind of test to be used naturally fell upon one of the series of performance tests described in "A Scale of Performance Tests," by Pintner and Paterson. The one that seemed to be the most generally useful was the Goddard Form Board. The Binet system was out of the question, since it has not been adapted for use in any of the numerous languages of India. This form of test depends so much on language responses that it would be hopeless to attempt to get reliable results by means of it before the questions had been carefully translated into the language of the subjects whom it was proposed to test, and proved by actual use. This would have been doubly difficult in the case of the children whom the writer tested, as they belonged to two different language groups. The use of performance tests avoids this difficulty. The Goddard Form Board was chosen as the most practicable test with which to begin the series because it has been so favorably spoken of by eminent psychologists in America. It also seemed to the writer to require the exercise of several kinds of mental ability. Not one of the more than 700 boys and girls tested had ever seen a form board, it is safe to assert. Few, if any, of them in all probability had ever handled blocks of wood or other material of different shapes, much less tried to fit them into holes of corresponding shapes. To be confronted with a board full of holes and a lot of blocks, and to be told to put the blocks into the holes as quickly as possible, was a new situation for each of those children. Thus it was well adapted to test their intelligence. At the same time there was nothing unreasonable in the test, so perfectly simple is it.

The method of giving the test was in general that described in the book by Pintner and Paterson already referred to. The board with the blocks in their respective holes was placed on a table or bench before the subject, who was told to watch closely what the examiner did. The latter then took the blocks out of the holes one by one, and laid them in the standard order in three rows by the side of the board at the subject's right hand. The subject was then told to put the blocks back in their right places. A native teacher was always present to make sure that the examiner's words were understood. In case of doubt he would repeat the command. Three trials were given, and the time and errors (if any) were recorded. The time of the fastest performance was regarded as the index of the subjects psycho-motor ability. In practice it was found best not to ask for speed at the first trial, as that tended to confuse him, and sometimes resulted in wild dashes at the board with little effort to avoid errors. A correct performance was the first thing aimed at. Before the second trial, however, the subject was told to put the blocks in as quickly as possible. Before the third he was urged to his utmost effort for greater speed.

The subjects examined were of both sexes. With a very few exceptions all were attending schools of average excellence. All were town-bred. Thus in the two groups respectively the home life and advantages, were practically the same.

The tests, with few exceptions, were given in the afternoon and generally in an unoccupied room or part of a hall in the school buildings, so that both subject and examiner might be as free as possible from distractions of any kind. Avoidance of distracting sounds was in many places impossible, as can be readily understood by any one who has visited even the best of schools in India, where the hot climate requires that doors and windows be kept wide open. As the pupils are used to these conditions, they need not be regarded as invalidating the tests.

Aside from the effort to secure an equal number of Brahman and Panchama children as subjects, no attempt at selection among the children of the schools where the tests took place was made. If this was done surreptitiously by teachers in any of the schools, it would tend to weed out the dullest ones, since the teachers would wish to have their pupils make as creditable a showing as possible. This would of course result in a higher average performance.

No performance time longer than 81 seconds was included in calculating the medians, in order to exclude records that

suggested decided mental backwardness, and to arrive at norms showing the average ability of ordinary school-going children in South India. Not more than one or two records were thus rejected in any one age group. These were only in the younger ages.

*Graph showing median Times of Pancharamas,  
Brahmans, Americans, from age 4 to 14*

Goddard Form Board

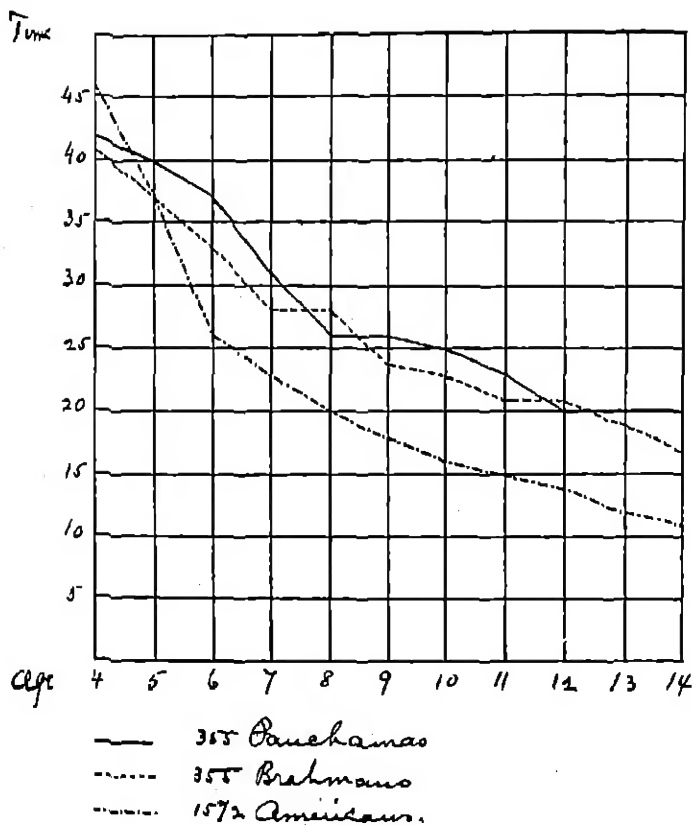




TABLE I—PANCHAMAS

Age	4	5	6	7	8	9	10	11	12	13	14
Time											
14											1
15									3	1	
16								1	2		2
17						1			3	3	1
18						2		5	5	5	4
19					3	1			5	4	2
20							2	3	3	5	3
21				1	1	1	2	3	4	5	2
22				1	3	2	5	3	3	3	2
23					2	2	4	1	3	2	3
24				1	5	3	1	2	3	1	
25				1	4	3	3	5	1	1	
26			1	5	2	3					1
27			1	4	2	4	3	5	1	2	
28		1	1	7			3	1			
29		1	1	3	2	1			1	1	1
30		1	2	5	1	1	3	3		2	
31		1	1	4	2	3					
32		1	1	3	2			1			
33			2	3	1						
34			2		2	1	1			1	
35		2		4		3					
36	1		4	2	1	1	1				
37	1		5	4	1	1					
38			1	2							
39			1	1	2						
40		2	2	3							
41		1	1		1	1	1	1			
42	3			1	1						
43		1	1								
44			1								
45		5	1								
46	1										
47		1									
48			2	1							
49	1	1									
50											
51			1								
52				1							
53											
54	1	1									
55			1								
56											
57			1								
58											
59											
60											
61			1								
62											
63											
64											
65				1	1						
66											
67	1										
68											
69											
70											
71											
72				1							
73											
79			1								
81		1									
Total	9	16	36	59	39	33	30	37	39	35	22
75 Percentile	37	32	33	28	24	23	22	19	18	18	18
Median	42	40	37	31	26	26	25	23	20	20	20
25 Percentile	49	45	43	37	33	30	28	27	22	23	22
Quartile	6	6.5	5	4.5	4.5	3.5	3	4	2	2.5	2

TABLE II—BRAHMANS

Age	4	5	6	7	8	9	10	11	12	13	14
Time											
14									1	1	6
15								1	2	2	2
16								1	2	2	2
17							1	1	1	6	4
18					1	1	2	8	1	7	6
19						1	2	5	7	7	2
20			2			3	2	2	4	5	
21				2	1	4	4	2	2	1	1
22				1	3	3	2	2	6	1	
23			1	4	1	3	2	4	3	1	
24			1	5	2	2	4	1	1	1	1
25	1		2		3	3	2	2	2	3	
26	1	3	1	4	3	1		5	1		
27		1	2	10	3	4	2	1			
28				5	4			1			
29				1	1	2	1	1	1		
30			3	5	5		2		1		
31		1		7						1	
32	1		4	2	1		1				
33		1	2	1	2						
34	1		1			1					
35			1	1	1		2				
36		1	2		1						
37		1		3	1						
38			3	1		1					
39		1	1	2		1					
40		1	1		1						
41	1		2	1							
42		1									
43			1		1						
44											
45	1	1			1						
46	2	2	2	1	1	1					
47		1									
48											
49			1								
50											
51		1	1								
52				1							
53											
54				1							
55											
56			1								
57											
58											
59				1							
60											
61											
62											
63											
64											
65											
66											
67	1										
68											
69											
70											
71			1								
72											
73	1										
79											
81											
Total	10	16	36	59	39	33	29	37	39	35	22
75 Percentile	32	27	27	26	24	21	20	18	19	17	15
Median	41	37	33	28	28	24	23	21	21	19	17
25 Percentile	46	45	41	32	32	27	25	25	23	20	18
Quartile	?	9	7	3	4	3	2.5	3.5	2	1.5	1.5

Tables are appended showing in detail the results of the tests.

For convenience of comparison a graph also has been prepared showing the median times of the Panchamas and Brahmans tested by the writer, to which has been added a line showing the median times of American children of corresponding ages according to Sylvester's tests. It will be noted that the Panchamas are from one to four seconds slower than the Brahmans at most ages. The average difference is about two and one-half seconds. Perhaps this can be accounted for by the great difference in the social and educational opportunities enjoyed by the two groups in the past, and by the difference in their environment, which any one acquainted with conditions in India can not fail to have observed.

When we compare the medians of the Brahman children with those of the American group, we find an even more striking difference. At the age of four the American are much slower than the Brahman children, the median times of the two groups being 46 and 41 seconds. At age five, however, the American children have caught up with the Brahmans, the medians for each group being 37. At six years the median time of the American group is 26 seconds, or seven seconds less than that of the Brahmans. In other words, between the ages of four and six the American children have increased the speed of their performance 20 seconds, while the Brahmans have increased theirs only eight. From this point onwards the Americans maintain their lead, and are from five to eight seconds faster at every succeeding age. The average difference is about six and one-third seconds.

The task of assigning causes for this very considerable difference is too large and complicated to attempt fully here. One fact is patent, viz., that the advance made by American children between the ages of four and six is much more rapid than that made by even the quickest group of Indian children of corresponding ages. It is generally accepted that children in the tropics mature earlier than in a colder climate. Presumptive evidence in support of this is found in the fact that the "age of consent," i. e., the age at which a girl may be legally married is fixed by law in India at twelve years. So it is not surprising that the Indian children at age four should be capable of a faster performance than the American four-year-olds.

Again, the gain of 20 seconds in speed of performance made in two years by the American children, as compared with a gain of only eight seconds made in the same period

by the Brahman group, may be accounted for at least in part by the better system and methods of education prevailing in the United States. The kindergarten age lies between four and six. India has not begun to approach America in the thoroughness with which the kindergarten has been developed, nor in the extent to which it has been employed.

Whether these considerations account fully for the differences noted above, or whether the large difference in the median times of the American and the Brahman groups from six to fourteen years of age indicates a fundamental difference in the mentality of the two groups, is a question that need not be discussed here.

One or two other facts of interest are to be noted. First, there is a check at age seven in the case of the Brahmans. They make no increase in speed from age seven to eight, the median times for the two ages being the same. Another occurs at age eleven. There is a similar check in the case of the Panchamas at age eight.

Again, the Panchamas show no increase in the speed of their performance after the age of twelve. So far as the form board shows, they have attained their full mental growth at this age. This, as of course all other inferences based on a single test, would need to be confirmed by other tests.

The quickest time made by any Indian child was 14 seconds, and was made by only three, viz., one Brahman of 12, one of 13, and one Panchama of 14 years. The fastest time made by any American child, as recorded by Sylvester, is nine seconds, and was made by children of 11, 13 and 14 years.

## A STATISTICAL METHOD FOR INCOMPLETE ORDER OF MERIT RATINGS

By M. J. REAM, Carnegie Institute of Technology

Rating scales are being used in an increasing number of industrial fields. The diverse situations in which ratings of employees are now being attempted have made necessary certain adaptations of the standard rating procedure. When a number of judges rate a group of individuals by placing them in an order of merit from the best to the poorest, direct comparisons are possible only in case each judge has rated the entire group. But if for some reason, such as lack of acquaintance, some of the judges were able to place in rank order only a portion of the group, the resulting rank positions would not be comparable. A method has been devised for combining such incomplete order of merit ratings in which the judges rate unequal numbers of individuals.

A case in point is the W. T. Grant Company which operates a chain of retail stores throughout the East. This company wished to get estimates of the comparative efficiency of its store managers, but because of their scattered geographical locations, no one judge could know all the men to be rated. Accordingly at a company convention, instructions for rating were given the judges as follows:

"Sheet I attached is a list of all the store managers of the company. First, go through the list and check the names of those with whom you have worked at least six months. Check any others with whom you feel sufficiently well acquainted to judge of their executive abilities. Copy the list of checked names on Sheet II.

"On each remaining sheet, a specific trait is listed and explained. Under this heading write the names you have checked according to merit in this particular trait. Put the name of the best man at the top and the poorest at the bottom. . . . ."

The results of this rating contained lists with as many as fifteen men ranked and lists with as few as three ranked. The problem was to combine the estimates of the judges to get the most probable order of merit in each trait.

A survey of the literature on the subject gave but one suggestion, an article by Professor E. L. Thorndike.<sup>1</sup>

---

<sup>1</sup> Thorndike, E. L. The technique of combining incomplete judgments of the relative positions of  $n$  facts made by  $n$  judges. *Journal of Philosophy, Psychology and Scientific Methods*, 1916, 13, 197-204.

He suggests that the first step is to get a rough approximation to the true order by mere inspection. The next step is to compare each individual with his neighbor in terms of the percentage of judges (using only those who rated the two individuals in question), who rated the former individual as lower numbered than the latter. Professor Thorndike gives a table in which any given division of opinion among 2, 3, 4, 5, 6, . . . . 15 judges may be expressed in terms of the median deviation. This process of comparison is continued with two or three next neighbors, and the several values determined from the table are averaged. Changes in position are made on a basis of these comparisons.

This method was tried with the Grant Company data but was far from satisfactory. Most laborious was the computation of the *average* differences in terms of the median deviation, from direct comparisons, next neighbor and second and third neighbor comparisons, and from indirect comparisons. After an average was obtained, a change of position was made on the basis of it, and new computations were necessary to justify the new position.

Furthermore, it was possible to get contradictory values by the method of indirect comparisons for individuals who were not directly compared by any judge. The indirect comparisons in addition to being laborious became confusing. Thorndike in his closing paragraph admits the possibility of ambiguity, ". . . . The resulting order should not without some consideration be used instead of the order derived from comparison with next neighbors only; for the next neighbor comparison, though not utilizing the data fully, is simple and unambiguous in its meaning."

To avoid these difficulties, a new method, or rather a modification of this method was devised and is presented as follows:

(1) For each rank-position on a list containing a given number of names, numerical values are assigned in terms of the standard deviation, assuming a normal distribution in each list of men rated. The values for any given rank-position are shown in Table I. For example, on a list containing six names, position one is assigned  $+1.49$ , position two  $+.68$ , position three  $+.22$ , position four  $-.22$ , position five  $-.68$ , and position six  $-1.49$ . These values were computed from Thorndike's Mental and Social Measurements, Table 22. The value  $1.49$  assigned to the first position on a list of six names represents the average amount in terms of  $\sigma$  of the upper 16.6% in the group, counted in from the extreme toward the average; the value  $.68$  assigned to position two represents the average

amount in terms of  $\sigma$  of the next highest 16.6% in the group. Values are similarly computed for other rank-positions. When a rank-position is reached which brings the percentage below the average, the  $\sigma$  value to be assigned becomes negative.

It is obvious that an individual ranked at the bottom of a list of three names should be assigned a different value from the individual who is ranked at the bottom of a list of twelve. Table II shows a sample method of tabulating the rating, in this case ten judges on thirteen individuals.

(2) On the basis of these assigned values, an algebraic average is found for each individual, and from these averages a new ranking is made with the individual with the highest average in position one, and the individual with the lowest average in position N.

(3) The verification and correction for this new order of merit is accomplished by the method of simple next neighbor comparisons only. By referring to Table II, the various rankings for any pair of individuals can be quickly obtained. For example, Subjects D and A have been given rank positions 9 and 10 respectively, on the basis of their numerical averages. In this case the two individuals appeared together on the lists of four judges. If Subject D was ranked higher than Subject A in three of the four lists, Subject D would be placed above regardless of the difference in numerical values previously assigned, which in this case was small. Subject D would then be compared with the next neighbor above until he proved to be a loser; similarly Subject A would be pushed down until he proved to be a winner. The simple direct comparison was made the crucial test in any change of position. In cases where the wins and losses were a tie, or in cases where no comparison of the two individuals was made, the numerical averages obtained in the second step above, were the criteria for determining relative positions.

It was found that at the upper and lower extremes of the order of merit, obtained in (2), practically no shifts were necessary as a result of the next-neighbor comparisons in (3), but frequent shifts were needed in the middle of the scale. However, the rank method of rating is generally considered less reliable in the middle of the scale than at the upper and lower ends.

The method just described proved to be much less laborious, and ambiguities and ties could be settled by reference to an independent numerical value. In the case of the Grant Company, the results of the rating differentiated quite strikingly the exceptionally good from the notoriously inefficient store manager.





TABLE I—Continued

TABLE OF STANDARD DEVIATIONS VALUES TO BE ASSIGNED FOR RANK POSITIONS

		Number of Individuals Rated											
		14	15	16	17	18	19	20	21	22	23	24	25
Rank	Rank	1	2	3	4	5	6	7	8	9	10	11	12
1	1.92	1.92	1.92	1.99	1.99	2.10	2.10	2.10	2.10	2.16	2.16	2.16	2.16
2	1.26	1.29	1.35	1.35	1.38	1.45	1.45	1.48	1.52	1.52	1.56	1.56	1.56
3	1.94	1.98	1.02	1.04	1.08	1.13	1.15	1.20	1.20	1.23	1.23	1.28	1.28
4	1.69	1.72	1.77	1.82	1.86	1.90	1.94	1.98	1.00	1.04	1.06	1.08	1.08
5	1.47	1.53	1.58	1.64	1.68	1.72	1.76	1.79	1.82	1.86	1.88	1.92	1.92
6	1.27	1.35	1.41	1.47	1.51	1.55	1.60	1.63	1.67	1.71	1.74	1.77	1.77
7	1.09	1.16	1.24	1.31	1.36	1.40	1.45	1.50	1.54	1.58	1.61	1.64	1.64
8	1.09	1.16	1.24	1.31	1.36	1.40	1.45	1.50	1.54	1.58	1.61	1.64	1.64
9	1.27	1.35	1.41	1.47	1.51	1.55	1.60	1.63	1.67	1.71	1.74	1.77	1.77
10	1.47	1.53	1.58	1.64	1.68	1.72	1.76	1.79	1.82	1.86	1.88	1.92	1.92
11	1.69	1.72	1.77	1.82	1.86	1.90	1.94	1.98	1.00	1.04	1.06	1.08	1.08
12	1.94	1.98	1.02	1.04	1.08	1.13	1.15	1.20	1.20	1.23	1.23	1.28	1.28
13	1.26	1.29	1.35	1.35	1.38	1.45	1.45	1.48	1.52	1.52	1.56	1.56	1.56
14	1.92	1.92	1.99	1.99	2.10	2.10	2.10	2.10	2.16	2.16	2.16	2.16	2.16
15	1.92	1.92	1.99	1.99	2.10	2.10	2.10	2.10	2.16	2.16	2.16	2.16	2.16
16	1.92	1.92	1.99	1.99	2.10	2.10	2.10	2.10	2.16	2.16	2.16	2.16	2.16
17	1.92	1.92	1.99	1.99	2.10	2.10	2.10	2.10	2.16	2.16	2.16	2.16	2.16
18	1.92	1.92	1.99	1.99	2.10	2.10	2.10	2.10	2.16	2.16	2.16	2.16	2.16
19	1.92	1.92	1.99	1.99	2.10	2.10	2.10	2.10	2.16	2.16	2.16	2.16	2.16
20	1.92	1.92	1.99	1.99	2.10	2.10	2.10	2.10	2.16	2.16	2.16	2.16	2.16
21	1.92	1.92	1.99	1.99	2.10	2.10	2.10	2.10	2.16	2.16	2.16	2.16	2.16
22	1.92	1.92	1.99	1.99	2.10	2.10	2.10	2.10	2.16	2.16	2.16	2.16	2.16
23	1.92	1.92	1.99	1.99	2.10	2.10	2.10	2.10	2.16	2.16	2.16	2.16	2.16
24	1.92	1.92	1.99	1.99	2.10	2.10	2.10	2.10	2.16	2.16	2.16	2.16	2.16
25	1.92	1.92	1.99	1.99	2.10	2.10	2.10	2.10	2.16	2.16	2.16	2.16	2.16

TABLE II  
RATINGS ON TRAIT X

	Subjects Rated												
	A	B	C	D	E	F	G	H	I	J	K	L	M
Judge No. 1 (8 comparisons)	-1.67	...	.91	-.91	.18	1.67	.....	-.18	.....	.....	-.50	.50	.....
Judge No. 2 (6 comparisons)	.....	.22	1.49	.....	.....	.68	-.68	.....	-1.49	.22	.....	.....	.....
Judge No. 3 (3 comparisons)	.....	0	.....	.....	.....	1.10	.....	.....	.....	.....	.....	.....	-1.10
Judge No. 4 (10 comparisons)	.....	-.39	1.76	.68	-.13	.13	.....	.39	-1.76	.....	-.68	1.04	.....
Judge No. 5 (5 comparisons)	.....	.....	.....	.....	.....	1.40	.....	0	-1.40	.....	.....	.53	-.53
Judge No. 6 (7 comparisons)	.....	1.59	.....	-.37	.....	.....	-1.59	.....	.....	.37	-.81	.81	0
Judge No. 7 (8 comparisons)	.....	-.50	.91	-.18	.....	.....	1.67	.18	-1.67	.....	-.91	.50	.....
Judge No. 8 (4 comparisons)	.....	1.27	.....	.....	.....	.32	.....	.....	-1.27	.....	.....	.....	-.32
Judge No. 9 (7 comparisons)	.....	-.37	1.59	.....	.0	.37	.....	.....	-1.59	.81	-.81	.....	.....
Judge No. 10 (9 comparisons)	.....	-.29	1.71	.....	-.60	.....	.....	.60	-1.71	.29	.....	.98	0
Total	-3.22	9.05	2.90	-2.01	.65	6.84	-4.29	.99	-10.89	1.25	-3.71	4.36	-1.95
No. times rated	5	8	4	5	3	6	4	5	7	4	5	6	5
Average	-.64	1.13	.72	-.40	.23	1.14	-1.07	.20	-1.55	.31	-.74	.73	-.39
Resulting rank order	10	2	4	9	6	1	12	7	13	5	11	3	8

# THE INTELLIGENCE OF CHINESE CHILDREN IN SAN FRANCISCO AND VICINITY

By KWOK TSUEN YEUNG, Stanford University<sup>1</sup>

## *The Problem*

The specific purposes of this investigation are:

1. To discover the general intelligence level of Chinese children in the vicinity of San Francisco by means of the Stanford Revision of the Binet tests;
2. To reveal sex differences, if such exist;
3. To compare the intelligence ratings of Chinese and American children of similar social status.

## *The Method of the Investigation*

The Stanford Revision of the Binet scale was given to 109 Chinese children during the spring of 1921. These children ranged in age from five to fourteen years and were all American born. Of the total number, 62 were boys and 47 were girls. In order to secure unselected groups, all of the Chinese children born in San Francisco or vicinity of the ages nine to ten and twelve to thirteen were tested. This first group numbered 49, the second group 33, the remaining 27 being of various ages from five to fourteen. Most of these children came from parents of the laboring and farming classes, chiefly immigrants from the region about the city of Canton. The parents had probably received education in China approximately equal to that secured by the typical village class of Chinese people. The present economic and social environment of the groups from which these Chinese children came probably resembles Chinese conditions much more closely than American life in the vicinity of San Francisco. The Chinese tongue, customs, beliefs, and manner of worship are all retained. They live in their own exclusive settlement, although the children attend the public schools and receive the same instruction as do the white children. The Chinese children master the English language readily.

In using the Binet tests, the vocabulary tests were omitted as unfair to the Chinese children, thus necessitating a small change in the scoring of the tests. Credit for each test passed was given as 2.4 months for the ages at which there are

<sup>1</sup> The writer is greatly indebted to Mr. G. M. Ruch, research assistant in education, Stanford University, for assistance in preparing this manuscript for the press.

six tests in the scale, and 3.4 months for each test at the ages having eight tests. Other than this change, the procedure was exactly that of the author of the tests.

In addition to the tests of intelligence, ratings of school work estimated on the seven-point scale were obtained from the teachers. The teachers, of course, were not given the test scores until after the ratings had been completed. The coefficient of correlation between tests and ratings was obtained by the Pearson product-moment formula. The Pearson coefficient of variability was also used in treating the data for

### *The Results*

TABLE I

THE DISTRIBUTION OF THE INTELLIGENCE QUOTIENTS OF THE 109 CHINESE CHILDREN OF AGES 5 TO 14 YEARS

I Q	Numbers	Percent
66-75.....	5	4.6
76-85.....	20	18.3
86-95.....	26	23.8
96-105.....	38	34.8
106-115.....	13	11.9
116-125.....	3	2.8
126-135.....	3	2.8
136-145.....	1	0.9
Total.....	109	

The median falls at 96.9, or approximately 97, and the empirical mode within the interval 96-105.

TABLE II

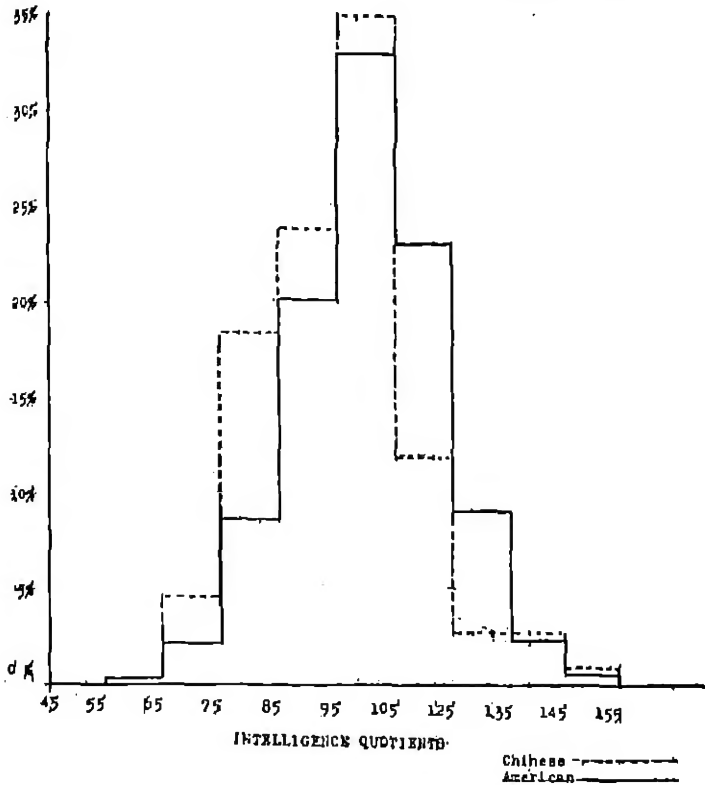
A COMPARISON OF THE I Q's OF THE CHINESE CHILDREN AND THE 905 UNSELECTED CHILDREN STUDIED BY Terman<sup>2</sup>

I Q	Chinese Children	American Children
56-65.....	0.0	0.33
66-75.....	4.6	2.3
76-85.....	18.3	8.6
86-95.....	23.8	20.1
96-105.....	34.8	33.9
106-115.....	11.9	23.1
116-125.....	2.8	9.0
126-135.....	2.8	2.3
136-145.....	0.9	0.55

The median for the unselected American group is 99 in comparison with 97 for the Chinese children studied. This difference as well as the somewhat more restricted range of intelligence quotients of the Chinese group cannot be held to be very significant when the smaller numbers of the Chinese children are taken into account. The possible difference in the amount of selection represented in the two racial groups should also be called to attention in arriving at conclusions from the data presented here.

<sup>2</sup>Terman, L. M.: 'The Measurement of Intelligence, page 40.

Graph I shows the data of Table II in graphic form.



The following table shows the results of Binet tests given to several racial groups in the vicinity of Stanford University.<sup>a</sup> Except for the Chinese groups, the children were all in kindergartens, who, as stated by Terman, usually show a higher score than do children of higher ages.

TABLE III

Race or Nationality	Numbers	Median I. Q.
Spanish.....	37	78
Portuguese.....	23	84
Italian.....	25	84
Chinese.....	109	97
North European.....	14	105
American.....	49	106

<sup>a</sup> Unpublished data.

It will be seen that the Chinese rank higher than the South Europeans and below the North Europeans and Americans. However, the Chinese children are older and such comparisons are probably not as unfavorable to the Chinese in comparison with the American children as seems to be the case at first glance.

In Table IV which follows, the distributions of the I Q's in percentages are given for the two unselected groups of Chinese children, i.e., the 9-10 year olds and the 12-13 years olds, in comparison with Terman's group of 905 unselected American children.

TABLE IV

THE DISTRIBUTIONS OF THE I Q IN PERCENTAGES FOR THE TWO UNSELECTED CHINESE GROUPS IN COMPARISON WITH UNSELECTED AMERICAN CHILDREN

I Q	Terman's 905 Unselected American Children	Unselected Group of 49 9-10 Yr. Old Chinese Children	Unselected Group of 33 12-13 Yr. Old Chinese Children
	%	%	%
56-65.....	0.33	0.0	0.0
66-75.....	2.3	4.1	6.1
76-85.....	8.6	12.2	21.0
86-95.....	20.1	26.5	21.0
96-105.....	33.9	51.0	30.3
106-115.....	23.1	6.1	6.1
116-125.....	9.0	0.0	3.0
126-135.....	2.3	0.0	9.0
136-145.....	0.55	0.0	3.0

The median I Q for the American, Chinese 9-10 year old, and Chinese 12-13 year old children falls, respectively, at 99.0, 97.4, and 96.0.

The following Table (V) shows the same two Chinese groups in comparison with unselected American groups of the same chronological ages.

The distributions for the Chinese children reveal tendencies for the I Q's to pile up at the average interval 96-105, and to thin out at the upper ranges, in comparison with American children of the same ages. The differences are not strikingly large and the two factors of selection and language familiarity should again be called to attention in interpreting the results.

TABLE V

COMPARISONS OF UNSELECTED CHINESE AND AMERICAN CHILDREN OF AGES 9-10 AND 12-13<sup>4</sup>

I Q	Ages Nine to Ten Years		Ages Twelve to Thirteen Years	
	American %	Chinese %	American %	Chinese %
66-75.....	2.0	4.1	5.0	6.1
76-85.....	12.0	12.2	15.0	21.0
86-95.....	21.0	26.0	20.0	21.0
96-105.....	38.0	51.0	28.0	30.3
106-115.....	6.0	6.2	19.5	6.1
116-125.....	1.0	0.0	11.0	3.0
126-135.....	1.0	0.0	8.0	9.0
136-145.....	0.0	0.0	0.0	3.0

*Sex Differences*

Table VI presents the data on sex differences for the total number of Chinese children tested.

TABLE VI

COMPARISONS OF THE SEX DIFFERENCES IN THE 109 CHINESE CHILDREN TESTED, IN WHOLE NUMBERS AND PERCENTAGES

I Q	Boys		Girls	
	Numbers	Per Cent.	Numbers	Per Cent.
56-65.....	0	0.0	0	0.0
66-75.....	4	6.4	1	2.1
76-85.....	15	24.2	5	10.6
86-95.....	16	25.8	10	21.3
96-105.....	16	25.8	21	44.7
106-115.....	7	11.4	6	12.7
116-125.....	1	1.6	2	4.2
126-135.....	3	4.8	1	2.2
136-145.....	0	0.0	1	2.2
Totals.....	62		47	

The sex differences are in favor of the girls, the medians being 93.5 for the boys and 99.9 for the girls. The distribution of the I Q's for the girls approaches the normal distribution much more closely than that for the boys, which is skewed downwards, particularly in the 76-85 interval. The extremes of the distributions for the two sexes are not strikingly dissimilar. The ratio between the coefficients of variability (Pearson) for the girls and boys is 0.63. The girls, then, are considerably less variable.

<sup>4</sup> Data on American children from Terman: Stanford Revision and Extension of the Binet-Simon Scale, pages 35, 37.

*Teachers' Ratings*

In order to make comparisons between the quality of school work and intelligence, ratings were secured from the teachers for the former on the basis of the 7-point scale. These ratings were correlated against the I Q's, giving a Pearson coefficient of  $0.52 \pm .048$ . The teachers' estimates correlated with school work (marks)  $0.46 \pm .060$ . These values are about the usual ones obtained by many investigators following the same method.

*School Progress*

A few facts concerning age-grade distributions of the Chinese children are of interest. The 109 pupils were distributed from the first through the eighth grade. Using the standard of ages 6 and 7 being normal for the first grade, ages 7 and 8 for the second grade, etc., it will be noted from the table which follows (Table VII) that the amount of retardation is greater for the Chinese than for the American children, although, with respect to acceleration the comparisons are more favorable to the Chinese. Table VIII gives similar facts when the age-grade classification is made upon the basis of the mental ages. The coefficient between C. A. and grade reached in school is  $0.84 \pm .028$ , and for M. A. and grade,  $0.88 \pm .015$ .

TABLE VII  
ACCELERATION AND RETARDATION FOR THE CHINESE AND AMERICAN GROUPS  
UPON THE BASIS OF CHRONOLOGICAL AGE IN PERCENTAGES

	Chinese	American
Accelerated		
One year.....	16.5	14.3
Two years.....	3.6	1.4
Three years.....	0.9	0.1
Total.....	21.0	15.8
Normal.....	25.0	44.3
Retarded		
One year.....	38.5	27.9
Two years.....	12.8	8.8
Three years.....	1.8	2.4
Four years.....	0.9	0.6
Total.....	53.0	39.7



TABLE VIII

ACCELERATION AND RETARDATION FOR THE CHINESE AND AMERICAN GROUPS UPON THE BASIS OF MENTAL AGE IN PERCENTAGES

	Chinese	American
Accelerated		
One year.....	23.8	15.6
Two years.....	3.6	3.2
Three years.....	0.9	0.4
Four years.....	0.9	0.1
Total.....	29.2	19.3
Normal.....	44.9	40.6
Retarded		
One year.....	19.2	27.2
Two years.....	5.5	10.2
Three years.....	0.9	1.9
Four years.....	.....	0.5
Total.....	25.6	39.6

Inspection of the distribution tables for age and grade will show that by the usual chronological age-grade classification, there is somewhat more acceleration and considerably more retardation among the Chinese. As would logically follow, when mental age is used, there is greater retardation for the American children. In other words, upon a mental age-grade basis the distribution of the Chinese children follows very much more closely the theoretically just basis than with the Americans. The excessive retardation of the Chinese children by the usual standards is very likely due to forced retention of children in grades below those indicated by the mental ages because of unfamiliarity with the English language.

#### *Occupational Status*

The occupations of the fathers of the Chinese subjects were obtained and classified upon the 5-fold basis recommended by Taussig. Upon this basis, the great majority of the Chinese children came from parents of classes 3, 4, and 5. Laundry workers and ranchers were most frequent. As has been repeatedly found with American children there is considerable correlation between the I Q and social status, this for the Chinese being  $0.49 \pm .049$ . The data indicate that the coefficient of correlation is lowered because large numbers of the Chinese children test in intelligence above their social and economic status.

## CONCLUSIONS

1. No striking differences in the intelligence of Chinese and American children are indicated by the results of this investigation. The Chinese children tested at about the level of Americans and North Europeans and markedly above South Europeans. The median I Q for the Chinese group was 97 in comparison with 99 as found by Terman for the 905 unselected American children studied.

2. Sex differences may be indicated by the data, although the numbers are entirely too small for generalizations. The median I Q for the girls was found to be 99.9 in comparison with 93.5 for boys.

3. Girls were only .63 as variable, according to the Pearson coefficient of variability.

4. Retardation is more frequent among the Chinese pupils than the American, although the Chinese are more perfectly classified in school grade with reference to mental ability than are the American pupils.

5. The Chinese of the vicinity of San Francisco belong chiefly to the lower levels of occupational status although there is a greater observable tendency for mental level to range above economic and social classification.

## METHODS FOR THE SELECTION OF COMPTOMETER OPERATORS AND STENOGRAPHERS

By M. A. BILLS, Carnegie Institute of Technology  
Communicated by C. S. YORUM

This study was made to determine if certain tests of the Bureau of Personnel Research of Carnegie Institute of Technology would serve first, as a basis for the selection of applicants for courses in stenography and comptometer operation in a technical night school; and second, to determine if from a group of applicants it was possible to select successful comptometer and stenographic operators without an extended try-out. That is, the first purpose demanded that failures be eliminated and the second that sure successes be selected.

Both purposes have been accomplished and the tests are to be used next fall as entrance requirements for the technical school, and as soon as employment is resumed<sup>1</sup> by the firm interested, as guidance in hiring comptometer and stenographic operators.

Both because of the problem presented and because of the methods of attack, which differ from those previously used, the study seems of special interest. The correlation between efficiency in the tests and efficiency in work has been the prime consideration in former studies.<sup>2</sup> In this study, because of the nature of the problem, correlations became merely a by-product, and critical scores the fundamental consideration. Also, a battery of tests has been used rather than a single test, with the idea in mind, that the more measures we could get of a person, the higher would be our ratio of success in selecting promising applicants.

The technical night school where the tests were given was in close connection with a large manufacturing firm and practically all of the teachers held positions in the firm and were therefore well able to judge the quality of work that would be acceptable in the business world. The ages of the students ranged from 16 to over 30. Practically every nationality was

<sup>1</sup> On account of the business depression, the firm is doing practically no employing at the present time.

<sup>2</sup> A Standardized Test for Office Clerks—L. L. Thurstone, *Journal of Applied Psychology*, 1919, 3,—pp. 248-251.

The Value of Mental Tests for Office Clerks—M. J. Ream—To appear soon in *Factory*.

represented and all classes of home conditions. All students understood English.

Three tests were given to the entire school (139 cases). The tests were:

1. A general intelligence test, designated in this article as Test VI.

2. A test devised to show special aptitude in stenography and typewriting, designated in this article as Test VIII. This test consists of five parts.

3. The Will-Temperament test, designated in this article as Test IX adapted from Downey's test and so revised that it can be used as a group test. This test has ten parts.

The criteria used were the failures and successes in the different classes. The teachers were asked not only to grade the student, but to indicate the line of work for which they considered her best adapted. A person was considered a failure when her teacher did not recommend that she continue work in her course or enter an advanced course. A successful student was one that was recommended to continue her work. For example, if the teacher of stenography recommended that a student take up typing only, that student was considered a failure. If the teacher in typing recommended that a student continue in typing or take up stenography, that student was considered a success, while she was considered a failure if clerical or messenger work was recommended.

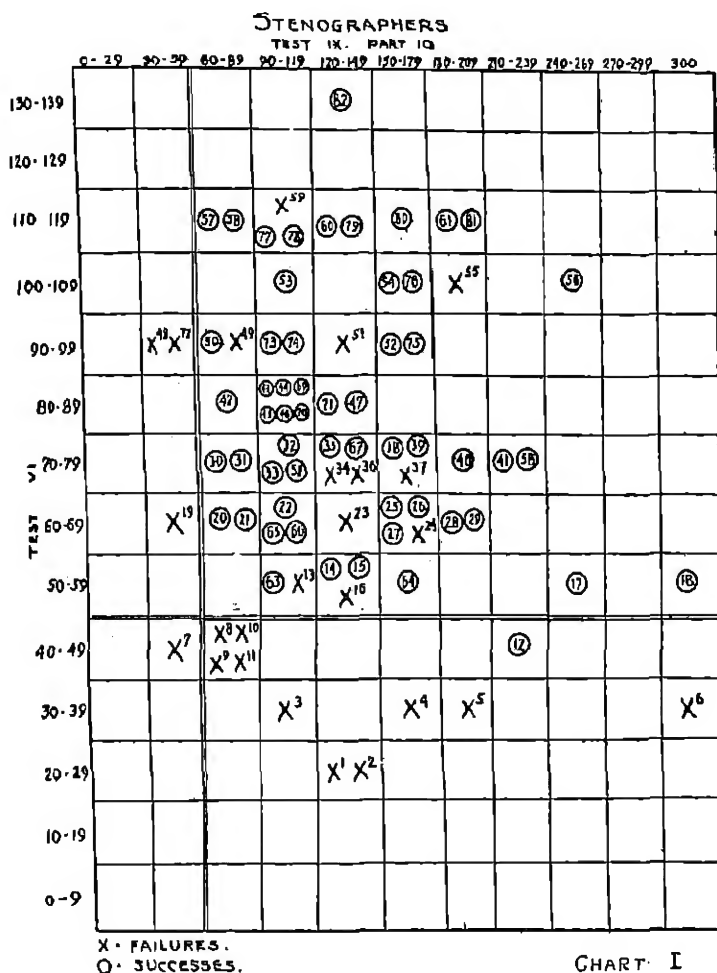
Of those tested 84 were in the course in stenography and 35 were in the course in comptometry. The remainder were in a preparatory course and will not be considered. Of the stenographers, 80 of the 84 took the entire series of tests and will be considered throughout the discussion. The other four have been included in the discussion of the tests which they took. Of the comptometer operators, the entire 35 took all three tests. Of the 84 stenographers, 26 were rated as failures. Of the 35 comptometer operators, 14 were rated as failures.

The tests that we found to be most useful in eliminating failures and picking successes were Test VI, Parts 1 and 5 of Test VIII and Parts 2, 3, and 10<sup>3</sup> of Test IX. As a whole it was found that Test VI (general intelligence) and Test IX (will-temperament) were the most successful in eliminating failures and that Test VIII (special ability) was the most efficient in selecting stenographers who were certain to be successful.

---

<sup>3</sup> These correspond to the parts of the will-profile that Miss Downey has called motor inhibition, speed of decision and perseverance.

Following are the results for stenographers in detail. We first attempted to eliminate failures. Our method can best



be illustrated by charts. In Chart I we have plotted scores made in Test VI against the scores made in Test IX, Part 10. We have indicated in the chart the successes and failures (O standing for successes—X standing for failures). Each X and each O stands for a person; each person has also been

## STENOGRAPHERS

## TEST IX. - SPEED OF DECISION OBJECTIVE - PART 3

TEST IX. MOTOR INHIBITION PART 2

	8-11	12-15	16-19	20-23	24-27	28-31	32-36
0-9				(63) (50) X <sup>59</sup>			(61)
10-19			(44) (37)	(81)	(31) (35) X <sup>72</sup> X <sup>23</sup>		
20-29		X <sup>16</sup> X <sup>68</sup> X <sup>83</sup>	(77) (60) (14) (47) (18)	(33) (96) (20)	(21) (45) (46)		(15) (65) X <sup>49</sup>
30-39	(47) (50)		(53) (26) X <sup>37</sup>		(70)	X <sup>24</sup>	
40-49	(17)	(34)	(23)	(43) (71)	(66) (64) X <sup>19</sup>		(57)
50-59	(62)	X <sup>4</sup> X <sup>1</sup>	X <sup>9</sup> X <sup>7</sup>		(72) (38) (40) (63)	X <sup>51</sup>	(29)
60-69	X <sup>13</sup> X <sup>3</sup>			X <sup>48</sup>	(79) X <sup>65</sup>	(67)	(27) (56)
70-79		X <sup>5</sup>	X <sup>11</sup>		(52) (16)		
80-89	X <sup>2</sup>	X <sup>10</sup>	X <sup>35</sup>	(19) (18)	(25) (84)	(29) (75)	(54) (12)
90-99				(18)	(82)		
100-109							
110-119					X <sup>6</sup>		

X - FAILURES

O - SUCCESSES

CHART II.

numbered. The numbers below 63 represent Juniors and the numbers 63 and above represent Seniors. It will be seen that 50 seems to be a critical score in Test VI. If we had refused admittance to all scoring below this point we would have eliminated 11 failures and only 1 success.<sup>4</sup> If we continue the elimination on the basis of a critical score of 60 in Part 10, Test IX, we eliminate 3 more failures and no successes. This makes a total of 14 failures eliminated out of the 26, and only 1 success. The critical scores are indicated by double lines on the charts. Now if we turn to Chart II, where we have plotted Part 3 of Test IX against Part 2, Test IX, we find that the combination of a score above 50<sup>5</sup> in Part 2 and below 20 in Part 3 indicates certain failure. This combination would eliminate 7 out of the 26 failures. The persons have been numbered the same as in Chart I. Comparing the two charts, we see that 5 of the 7 have already been eliminated with Test VI, so that, considering the two charts, we find that we can eliminate 16 out of the 26 failures in stenography.

Putting this into a four fold table, we have:

Failed in tests and failed in course.....	16 cases
Failed in tests and succeeded in course.....	10 "
Succeeded in tests and succeeded in course....	56 "
Succeeded in tests and failed in course.....	1 "
Total .....	83 "

This means that 86% were correctly placed by preliminary examination. It is also important to note in this connection that no person who has become a Senior received a score in Test VI of less than 50, in Test IX, Part 3 of less than 24, and in Test IX, Part 2 of more than 50.

On account of the complication involved in the use of four tests as part of the entrance requirement, we have recommended to the technical school that two tests only be used in the elimination of those from the course who are certain to fail. These tests are Test VI and Part 10 of Test IX. Others are to be given however, for their scientific interest.

<sup>4</sup> Upon personal discussion with the teacher we found that this one success was a doubtful one. The probability that this elimination can have taken place by mere chance is 1 in 26,518 worked out by the ordinary theory of probability, so that we may be reasonably sure of the causal relation between Test VI and success in stenography.

<sup>5</sup> A high score in Part 2 indicates a poor standing.

## STENOGRAPHERS

## TEST 8 PART 1

	2-39	4-59	6-79	8-99	10-119	12-139	14-159	16-179	18-199	20-219
200-209					(34)			(76)		
190-199								(61)		
180-189			X <sup>55</sup>			(22)			(78)	
170-179			(77)(35)	X <sup>48</sup>	(31)		(25)(33)		(50)(62)	
160-169		(14)	(33)(43) X <sup>36</sup>	X <sup>23</sup>	(45)(39)		(29)	(47)	(58)	
150-159		(52)	(67)	(34)	(60)	(47)(13)				
140-149						(56)(15) (28)(71)	(61)	(80)(75)	(57)	
130-139		(17) X <sup>7</sup>	(64) X <sup>2</sup> X <sup>40</sup>	(38)		(69)	(74)			
120-129	(70)	X <sup>37</sup> X <sup>5</sup> (70)	(12) X <sup>9</sup>	(10)(44) (18)(46)(11) (2)(66)	(27)(13) X <sup>24</sup>	(32)(69)	X <sup>51</sup>		(79)	
110-119	X <sup>13</sup>	(15) X <sup>1</sup> X <sup>9</sup>		(43)	(23)					
100-109	X <sup>10</sup>	X <sup>11</sup> X <sup>8</sup>	X <sup>3</sup>	(20)						
90-99	(30)					(63)				
80-89										
70-79										
60-69							X <sup>12</sup>			
TEST NOT TAKEN	X <sup>1</sup> X <sup>11</sup> X <sup>48</sup>	X <sup>83</sup>								

X = FAILURES  
O = SUCCESSSES

CHART III.



Second, we attempted to select those from the group who were certain of success. In Chart III we have plotted Part 1 of Test VIII and Part 5 of Test VIII against each other. Again, O's represent successes and X's failures. Each person is numbered as before. The numbers below 63 represent Juniors, and 63 and above represent Seniors. It will be seen that a score above 10 in Part 1 and above 130 in Part 5 indicates certain success. These grades should be used as critical scores in the employing of applicants when a large number are present from whom to make the choice and when no chances are to be taken on the success and failure of the applicant. In every employment office such instances arise. Considering Chart III from a mathematical view point, we find that each part predicts successes and failures as follows:

*Part 5—Critical Score, 130*

Failed in test and failed in course.....	15	cases
Failed in test and succeeded in course.....	19	"
Succeeded in test and succeeded in course.....	39	"
Succeeded in test and failed in course.....	7	"
Total .....	80	"

This shows a correct choice of 67%.

*Part 1—Critical Score, 10*

Failed in test and failed in course.....	23	cases
Failed in test and succeeded in course.....	23	"
Succeeded in test and succeeded in course.....	35	"
Succeeded in test and failed in course.....	3	"
Total .....	84	"

This shows a correct choice of 69%.

A combination of the two parts of Test VIII gives the following results:

Failed in test and failed in course.....	26	cases
Failed in test and succeeded in course.....	30	"
Succeeded in test and succeeded in course.....	28	"
Succeeded in test and failed in course.....	None	
Total .....	84	"

This gives 64% correct choice. As will be noted, this correct choice is not as great as for Test VI but the important thing to note here is that success in the tests predicts success in the course perfectly (100%).

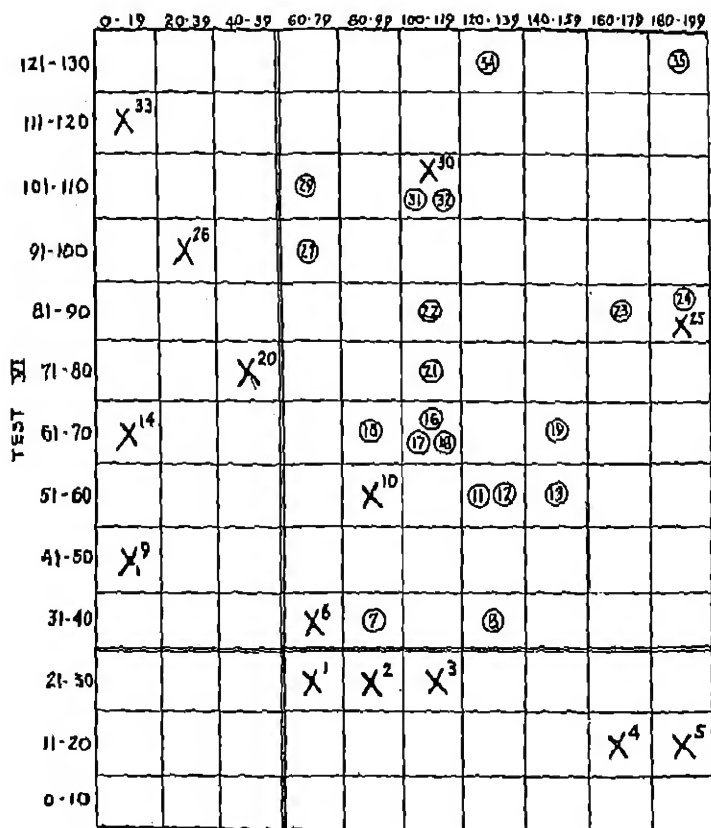
If, on the other hand, we take 8 as the critical score in Part 1 and 120 as the critical score in Part 5, we have:

Failed in test and failed in course.....	19	cases
Failed in test and succeeded in course.....	14	"
Succeeded in test and succeeded in course.....	44	"
Succeeded in test and failed in course.....	7	"

Total ..... 84 "

## Comptometer Operators

### TEST IX. PART 10



X • FAILURES  
O • SUCCESSES

CHART IV

This shows a correct choice of 77%.

In picking our applicants that are sure of success we have, however eliminated some who would have succeeded, and therefore this criterion should only be used when the number of applicants considerably exceeds the openings in the company; or when the selecting of a failure would be a very serious matter.

In Chart IV, we have plotted the scores of the comptometer operators in Test VI and Part 10 of Test IX. It will be seen that a combination of the two tests will eliminate 10 of the 14 failures. The critical scores are 30 for Test VI and 60 for Part 10 of Test IX. The chance that this is a mere chance elimination is 1 in 25,921. We have found no combination of tests for the comptometer operators that would pick successes more favorably than these two tests with which we have eliminated the failures. This is due to the fact that at the present time we have no special ability test for comptometry. We hope in the near future to work out such a test.

Putting the results on comptometer operators in mathematical form, we find that they predict success and failure as follows:

Failed in tests and failed in course.....	10 cases
Failed in tests and succeeded in course.....	0 "
Succeeded in tests and succeeded in course....	20 "
Succeeded in tests and failed in course.....	5 "
Total .....	35 "

This shows a correct choice of 85%.

### RESULTS

I. A battery of tests is more effective both in eliminating failures and picking successes, than any single test.

II. Of the single tests, General Intelligence is the most efficient for eliminating failures.

III. Of single tests, the Special Ability test is most efficient for selecting successes.

IV. Failures can be predicted by the tests with over 85% accuracy.

V. Successful stenographers can be selected.

## MINOR STUDIES IN EDUCATIONAL PSYCHOLOGY

By FRANCES I. GAW, Boston Psychopathic Hospital, Boston, Mass.

### I. SCHOOL RATINGS AND MOVING PICTURES

The educational value of moving pictures and their influence upon behavior have been much discussed questions in the last few years. Educational films of various sorts have become common, and the release of such a picture as the "Fit to Fight Film" in many camps during the war gave official recognition of the educational possibilities of movies. Dr. William Healy in his "Individual Delinquent" in reporting Juvenile Court cases where movies have directly affected conduct, says. "In nearly all people visual memory and visual imagery play the most dynamic part in mental life. . . . When it comes to motion pictures we have added elements of force for the production of either good or bad." The importance of moving pictures as a method of popular instruction is borne out by the existence of such a publication as the Educational Film Magazine.

This work was done to study the influence of movies on the conduct and school ratings of 337 children in a suburb of Boston. In this community there are no moving picture theaters, and the sentiment against them is strong. The children studied were in the 6th, 7th and 8th grades. They were given a questionnaire to determine how many movies they saw a month. Their ratings in studies and deportment were supplied by their school teachers, and correlations were then calculated between these averages and the number of movies attended monthly. The coefficient was found to be, in the case of school studies and movies, —13. The correlation though inverse, is so low that practically no relationship is shown between the two factors considered. In the case of deportment grades and movies attended, the situation is similar with a correlation of —12. The average number of movies attended monthly was .97, or slightly less than one a month. Only 44 children went as often as once a week, and only 6 more than once a week. Over half of the 337 children considered did not go to the movies at all. The two pupils with the lowest averages in their studies were among those who went to movies the most, although many of the other pupils low in their studies and those lowest in deportment went to the movies either an average number of times or not at all. A decided majority of the children with the best averages in both studies and deportment did not go to the movies at all.

The following questionnaire (for which the writer was not responsible) was used in gathering the data for this study:

- "1. Do you go to the movies regularly?
2. Do you go to the movies as often as once a week?
3. Do you go to the movies as often as once in two weeks?
4. Do you go to the movies as often as once in a month?
5. Do you go to the movies less frequently than once a month?"

Some children in answering became confused, contradicted themselves and in a few cases added a question. There were 4 or 5 questionnaires so inaccurate that they could not be counted in the correlation. Such a questionnaire might well be modified to read:

1. Do you go to the movies at all?
2. Do you go to the movies as often as once a month?
3. Do you go to the movies as often as once every two weeks?
4. Do you go to the movies as often as once a week?
5. Do you go to the movies more than once a week?

Whether or not this community is typical in regard to the number of movies the children attend, it is obvious that in this study the average movie attendance—once a month—is practically unrelated to either the school work or deportment of the children considered.

## II. RELATIONS OF STANFORD TESTS AND DEARBORN MAZE TESTS

In this study, the relationship was studied between the Stanford-Binet mental ages of a group of patients at the Boston Psychopathic Hospital and their scores on Test No. 7, General Examination 5 of the Dearborn Group Tests of Intelligence. The Dearborn test used consists of four mazes. It is explained to the subject that the first two are maps of a city which he is to traverse in the shortest way possible, keeping in the streets throughout. The last two are also maps of a city which he is to traverse, stopping at certain points, but he must follow two definite traffic rules: first, he can turn only to the right, and, second, he must always stop on the right-hand side of the street. The third maze is divided into four parts. The first one of these was used as an illustration in the cases examined. Since this study was made, however, the third part also has been used by the author of the test as an illustration. The two more difficult mazes of the test are probably made considerably easier by this change.

In addition to the patients tested, the Dearborn test was given to 36 normal adults connected with the Psychopathic Hospital. Most of these were physicians, social workers and psychologists. Although this is a group test, it was given individually to many of the persons tested, both patients and normal adults.

The total number of patients tested was 77. Seven correlations were calculated by the ordinary Pearson formula. (The character of the data rendered rank difference methods impracticable.) First, the Stanford-Binet and Dearborn scores of the entire group were correlated. The  $r$  was found to be .52. The cases were then grouped according to diagnosis. It should be noted here that no patients were tested who were obviously psychotic, or whose scores would probably be lowered by their condition at the time of the test. Seventeen fell into diagnostic groups which were too small to be dealt with separately. The correlations between Stanford-Binet scores and Dearborn scores of the groups large enough to use were as follows. Using 14 cases diagnosed psychopathic personality  $r$  was .36. When the records of two of the patients in this group were omitted, because of having driven cars under the same traffic rules as those in the third and fourth mazes of the Dearborn test, the correlation of the remaining twelve patients was .67. Fourteen cases diagnosed mental retardation (subnormal intelligence) gave a correlation of —.62. In this group, there were four cases of life ages respectively 15-4, 14-9, 12-11, 10-6. These 14 combined with a group of 8 cases diagnosed mental retardation plus conduct disorder, gave a correlation of —.17. A group of 13 cases of conduct disorder gave a correlation of —.15. When these cases were combined with 7 others diagnosed conduct disorder, plus mental retardation, the correlation found was 7.

In the group of normal adults the scores were consistently higher throughout than in any of the groups of patients. The accompanying

distribution compares the Dearborn scores of the normal adults, the cases of psychopathic personality and a group of the adult patients tested whose mental ages fall below 13 years, thirteen being considered the lower limit of average adult intelligence. Three cases, one at score 15, one at score 19 and one at score 23, refer to the 3 individuals (2 patients and 1 physician) who have driven cars in cities under the traffic rules used in the third and fourth mazes. The median Dearborn score for the normal adults is 12.5, for the cases of psychopathic personality 5.5, and for the patients with mental age below thirteen, 3.3.

Dearborn Score	0	1	2	3	4	5	6	7	8	9	10	11	12
Normal Adults.....	0	0	0	1	1	1	1	1	2	6	3	1	2
Psychopathic Personalities...	1	0	2	2	2	0	0	2	0	0	1	0	0
Patient with Mental Ages below 13.....	3	2	6	6	8	3	0	0	0	0	0	0	0

Dearborn Score	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Normal Adults.....	2	2	2	1	0	2	0	0	0	2	3	0	0	3
Psychopathic Personalities	0	1	1	0	1	0	1	0	0	0	0	0	0	0
Patient with Mental Ages below 13.....	0	0	0	0	0	0	0	0	0	0	0	0	0	0

While the ability concerned in the maze tests is seen usually to go hand in hand with the more general intelligence criterion, it at times plays a more special rôle, notably below or above the general intelligence rating. Among the cases under diagnosis of mental retardation, this is so frequent as to give the expression of relationship its markedly negative value,  $-.62$ . This is evidence of more disparate development of intelligence functions in this class of cases.

## COMMUNICATIONS AND NOTES

### INDUSTRIAL PSYCHOLOGY IN EUROPE

*To the Editors of the Journal of Applied Psychology:*

Sirs: I have just returned from a visit to the laboratories of industrial psychology in Europe, and inasmuch as several colleagues have suggested that a report of my observations would be of interest to the readers of the JOURNAL OF APPLIED PSYCHOLOGY, I take pleasure in submitting an informal résumé of my tour.

In all the countries visited, I found the applied psychologists interested in this field, though stressing divers points in approaching it.

In England, industrial psychology has a sort of official status, one important branch being under the auspices of the Industrial Fatigue Research Board (something like our National Research Council), which has been for several years investigating the relation between fatigue and various working conditions. Some publications of the Board are already available, a list of which may be secured upon application to the secretary at 6 John St., Adelphi, London, W. C. 2. The nature of the current investigations may be denoted by one which Miss Smith is conducting upon the conditions under which women work in laundries. I was struck with special force by the novelty of some of the methods that have been devised for measuring fatigue—not, be it said, by means of the oft-tried mechanical instruments; the attempt is made, rather, to express it as a function of the job itself.

Another organization that indicates the English interest in industrial psychology is the National Institute of Industrial Psychology, 329 High Holborn, London.

Though the operating center of these two organizations is London, the animating center is largely Cambridge. Dr. Myers has been indefatigable in his efforts to inaugurate and secure support for the work. As part of his program he called Mr. B. Muscio from Sydney to lecture at Cambridge upon industrial psychology and to conduct investigations in the laboratory.

As I shall presently show by comparison, the English approach emphasizes one point of view particularly. For example, tests do not bulk as large there as in some other centers. Nevertheless I am very optimistic regarding the results of the English approach to the problem. It has bearings beyond its ostensible aim of ameliorating conditions of work. It constitutes a long step in the direction of the too-long neglected analysis of the job, and as such it may furnish information useful in vocational selection and guidance, in increasing output, and in many other economic functions.<sup>1</sup>

In France I found lively interest in the psychological problems of industry, though it is scattered among several circles. A great part exists as *le taylorisme*, which has developed into a veritable cult among French engineers. Its vogue has penetrated even into conservative university circles, as evidenced by the course given at the Sorbonne by M. Charles Cestre, upon American methods of industrial management. M. Cestre introduced me to a half dozen of the leading economists

---

<sup>1</sup> See H. D. Kitson, "Scientific Method in Job Analysis," *J. of Pol. Econ.*, XXIX (June, 1921), 508-514.

and engineers who have organized for the sake of giving impetus to such movements, the *Société d'encouragement au commerce et à l'industrie*, 4 rue de Rennes, Paris.

The chief investigations emanating from orthodox psychological sources are those carried on by M. J. M. Lahy at Paris. He has written a book, "*Le System Taylor et la physiologie du travail professionnel*." He is also secretary of the 1921 conference of industrial psychology which I shall mention presently.

Further French interest in industrial problems is revealed in several physiological studies, some upon fatigue, and some upon other effects and conditions of work. There is still another angle that is represented—the pedagogical. Here the chief interest centers around vocational guidance, with considerable thought for the *multitude* of war.

My visiting in Germany was confined exclusively to Berlin where I was conducted about chiefly by Dr. Piorkowski who, with Dr. Moede, edits the journal, *Praktische Psychologie*.<sup>2</sup> These investigators have an Institute with headquarters at the *Technische Hochschule* located in the Charlottenburg section of Greater Berlin. Their activities are manifold. At the *Hochschule* they administer tests to students and apprentices; in the capacity of consultants they have introduced occupational tests into several factories in Berlin; and they publish the journal.

The kinds of tests used follow largely the Muensterberg tradition of attempting to reproduce the work to be performed, in diminutive and elemental form. For example, the test for aptitude for running a machine requiring retention and reproduction of a certain pressure just made by the hand, consists of an apparatus upon which the subject can exert a pressure which is recorded, and which he afterwards reproduces as closely as possible. The two kymographic records are then compared, giving an "aptitude index." Further examples of the tests employed and the methods employed in evaluating them may be found in the pages of "*Praktische Psychologie*."

In order to save me time and energy in viewing all their activities, my German hosts had prepared a three-reel film showing the operation of many varieties of tests. This they presented to me in private demonstration. They informed me that they would be glad to distribute this in the United States at the price of three hundred dollars.

During the week I was in Berlin a series of intelligence tests was being administered to children just emerging from the *Mittelschule*, in order to select a certain number as entrants in a school for gifted children. A large group had already been selected on the basis of scholarship; and they were being still further sifted by means of intelligence tests. The day I observed the work about 150 girls were gathered in the grand council-chamber of the *Berliner Rathaus* for testing. The tests employed were the traditional word-building, computation, Ebbinghaus completion, and the like. The testing went very smoothly but very slowly. The children wrote on separate sheets of paper torn from booklets similar to the examination "blue-books" of American universities. For three hours that morning the children sat and took tests with only a fifteen minute intermission. The work of that morning represented only a small portion of the total undertaking, an entire week being required for testing all the children. The work of scoring papers must have been mountainous. Some of our Ameri-

<sup>2</sup> See H. C. Link, "A New Journal of Practical Psychology," *J. of App. Psychol.*, V (March, 1921), 85-88.



can psychologists who are specializing in forms of intelligence tests and in devices for increasing efficiency in their administration may find a good market for their wares in Berlin.

At Geneva, I found the *Institut J. J. Rousseau* likewise interested in industrial psychology. This institution, founded in 1912 under the direction of Professor Claparède, is devoted chiefly to the application of psychology to educational problems. But in the light of recent movements in vocational education, the Institute feels itself bound to direct a moiety of its energies into the industrial field. As early as 1916 a rather pretentious attempt was made to formulate a series of tests for telephone operators. The results were published in the *Archives de Psychologie* and furnished a pattern for several similar investigations in Europe. Occupations that the Institute has since approached analytically in various Swiss factories and technical schools are watchmaking, various operations of the machinist's trade, and several occupations for women. The report of this latter group of studies, M. Bovet, Director of the Institute, assured me, is in fair way to publication.

A considerable amount of attention is being paid on the continent to vocational guidance, or "*orientation professionnelle*," as it is called. I found two books upon the subject in French, and two journals—one in French, and one in Spanish. Germany has a national association of vocational guidance of which Dr. Moede is president. And all Europe has inaugurated a series of annual conferences. The first one was held last year at Geneva. The second (*Deuxième Conférence Internationale de Psychotechnique Appliquée à l'orientation professionnelle et à l'organisation du travail*), will take place at Barcelona September 28-30, 1921. The program is divided into three sections: Section 1, (a) Vocational guidance; (b) motion study; (c) occupational fatigue. Section 2, Psychographic methods (especially for schools). Section 3, Tests. The conference will be entertained by the personnel of *L'Institut d'orientation professionnelle* of Barcelona. I should like to have stayed over for this conference and have brought back a complete report of it but I had not made arrangements to miss the opening weeks of the university.<sup>3</sup>

From observations such as those I made, one may conclude that the interest in industrial psychology is just as keen in England, France, Germany and Switzerland, as it is in the United States. It is very intense, and is animated by high moral and patriotic motives of improving conditions of labor, raising the efficiency of production, guiding young people (especially those maimed in the war) into appropriate occupations, and educating them with effectiveness and dispatch. In each of the countries emphasis is placed upon one or another of these phases. In England, perhaps it is chiefly directed toward the amelioration of working conditions; in France, toward raising general technical efficiency; in Switzerland toward vocational guidance, or at least, vocational education.

Intense though the interest is, it is not as extensive, however, as that in the United States. (There were only about twenty in attendance at the last meeting of applied psychologists at Geneva.) One receives the impression that European psychologists look with considerable admiration upon the achievements of American psychologists in

<sup>3</sup> The editors are pleased to announce that through Mr. Kitson's efforts they hope to present the outstanding features of the conference in a later issue.

this field, particularly upon the personnel activities of the army. (I think I am betraying no professional or military secrets in remarking that there has been serious talk of adopting our personnel methods in the Swiss army.)

The European investigators are generally acquainted with our books upon the subject, but not with our periodical literature. Nowhere did I find in use the recent forms of our twenty-seven varieties of intelligence tests. Nowhere did I find test-booklets, "cross-out" elements, scoring stencils, simplified statistical methods and other aids to efficiency so common in the States. I suggest, therefore, that any American mental testers who are anxious to spread the use of their tests may profitably place on their mailing-lists the following European names:

Dr. C. S. Myers, Psychological Laboratory, Cambridge University.

Mr. B. Muscio, Psychological Laboratory, Cambridge University.

Mr. G. H. Miles, 329 High Holborn, London, W. C.

Mr. D. R. Wilson, 6 John St., Adelphi, London, W. C. 2.

Mr. Eric Farmer, 6 John St., Adelphi, London, W. C. 2.

Miss May Smith, 6 John St., Adelphi, London, W. C. 2.

Dr. W. Moede, Berlin, W. 30, Luitpoldstrasse, 14.

Dr. C. Florkowski, W. 30, Luitpoldstrasse, 14.

Professor Ed. Claparède, Psychological Laboratory, University of Geneva.

M. Heinis, Psychological Laboratory, University of Geneva.

M. Pierre Bovet, Director Institute J. J. Rousseau, Geneva.

Madame Chavennes, Institute J. J. Rousseau, Geneva.

Mlle. Chavennes, Institute J. J. Rousseau, Geneva.

M. J. M. Lahy, 12 rue Linné, Paris.

The trip showed me that the problems which confront industrial psychologists are similar, whether in Europe or America, and that when approached from the scientific point of view, they lead over quite similar pathways. It also showed me that in spite of the mere three thousand miles that separate our coasts there is regrettable lack of acquaintance between applied psychologists in Europe and America. And that in order for a maximum of progress to be made by both groups there must be deliberate and constant efforts towards facilitating a *rapprochement*.

Professor of Psychology, Indiana University,  
Bloomington, Indiana.

HARRY DEXTER KITSON

## THE NATIONAL INSTITUTE OF INDUSTRIAL PSYCHOLOGY OF THE UNITED KINGDOM

For some time the more progressive firms in this country have been interested in the application of psychological and physiological principles to industry and commerce. The Industrial Fatigue Research Board which was established in 1915 has already done much to show the possibilities in this field. Its activities are however necessarily confined to broad surveys and investigations of common problems and it was felt by many firms interested, that much of the value of its work would be wasted unless assistance could be obtained in the application of the principles to the special needs of individual firms. This work is obviously beyond the scope of a Board supported by public funds and in order to meet this need a National Institute of Industrial Psychology has been formed. This was incorporated under the Com-

panies Act on February 11, 1921, as an Association for Scientific Research. It has the support of a number of well-known business men and in addition has a Scientific Advisory Committee consisting of the heads of psychological, physiological and educational departments of Universities throughout the Kingdom who are interested in the practical application of the sciences of Psychology and Physiology. Private individuals have come forward with generous financial support, and in addition the Carnegie United Kingdom Trust is giving substantial help. Where investigations are carried out which are likely to result in an immediate advantage to a firm, fees are charged which cover the salaries of the investigators and include overhead charges of the central body.

The Institute is under the direction of Dr. C. S. Myers, M. D., Sc. D., F. R. S., and by arrangement with the Industrial Fatigue Research Board the services of their investigators may be utilized when required to supplement the work of the Institute's investigators.

On the industrial side of the Institute's activities investigations are now being conducted in factories concerned in the work of presses, box and can making, soldering, chocolate dipping and packing; interesting problems concerning layout, size of material, rest pauses, selection tests, seating accommodation and ventilation have arisen and are being examined.

On the commercial side the formation and trying out of a suitable series of selection tests for the clerical staff of large firms is being undertaken, which promises to prove most useful.

When more detailed information is obtained as a result of these and other investigations, the Institute will be in possession of material invaluable for Vocational Guidance. By cooperation with a number of leading education authorities definite and scientific lines of approach will be formed between the school and business life so that a boy or girl, on leaving school, will have the opportunity of obtaining more exact and scientific guidance than has been possible in the past. In this way it is hoped that much of the waste of national energy which takes place in the intermediate stage by haphazard drift and false starts will be avoided.

The Institute is therefore making a more widespread appeal to all those who are interested in this work. That it has been possible to advance so far in a time of such severe financial depression is a hopeful promise of greater activity in brighter times. Meanwhile active propaganda work is being carried on, both amongst employers and employees, and lectures dealing with the subject are being arranged. The results of the work and a full survey of the work of similar character carried out by other similar bodies abroad will be published by the Institute in the form of a quarterly journal.

The Secretary of the Institute is Dr. G. H. Miles and the head offices are at 329 High Holborn, London, W. C. 1.

---

We are indebted for the above account of the work of the National Institute of Industrial Psychology of the United Kingdom to the Director, Dr. Charles S. Myers.

## BOOK REVIEWS

KEMPF, EDWARD J. *Psychopathology*. St. Louis, C. V. Mosby Co., 1920, pp. 762.

"It is comparatively simple and very easy to be an extremist in anything, but it requires eternal care, sound common sense, and no little patience and endurance to maintain a progressively refining, healthfully constructive attitude toward the fundamental needs and pleasures of human nature." When the above statement appears as early as the third page of this recent contribution to psychopathological literature, the reader is inclined to assume a rather comfortable postural tonus and approach the volume with the expectation that perhaps at last has been found the long anticipated volume which is to safely point a course between the extremes of too rigid conservatism and too liberal psychoanalytic deductions. Surely a writer who voices the need of "eternal care" and "sound common sense" will not fail to heed his own warning. For the reader expecting a conservative interpretation of human adjustment, there is increasing disappointment. He will find here outlined, no medium course, but still another extreme contribution to psychopathological literature, which he must appraise and, after selection and rejection, weave into his own hypotheses.

Of one thing there can be no doubt. Kempf himself is sincerely convinced of his theory and is most eager to proselyte. In fact, his enthusiasm is so contagious that there is danger of uncritically accepting statements for which reason forces, if not rejection, at least serious question. In actual presentation of data, Kempf seems so anxious to get his view point before the public, that he takes no time to refine. As a result there is much to be criticized in his style. It is neither consistently popular nor consistently scientific. One finds tedious repetition, unnecessarily long and involved sentences, an occasional incomplete sentence and illustrations ad infinitum. It is to be regretted that "Psychopathology" is not presented with the fineness of expression of "The Yellow Jacket." Unkindly criticism might characterize Kempf's present style as he does his patients. "The large number of unique words which he used to express a simple thought was unusually impressive." (Page 436.) Again, "he used numerous polysyllabic words with little regard to their meaning." (Page 530.) It would indeed take a clever reader to grasp the meanings of "goodness" and "happiness" from a study of the wordy definitions on page 118. However, it is his coarseness of expression which can least be forgiven. Kempf himself calls attention (page 3) to the biological destructiveness of the tendencies to vulgarity and to prudishness. Certainly the way to overcome "prejudice, prudery, or mock finery" is not by actual coarseness of expression. The latter is quite unnecessary, however "delicate" the situation discussed.

To understand Kempf's dynamic, autonomic mechanism of the personality, one must appreciate first of all the tremendous emphasis placed upon the autonomic system. He assumes that "the different segments of the autonomic apparatus are stimulated to assume different types of postural tensions and activities, which give rise to an affective nervous stream, which, in turn, coordinates the projicient apparatus and compels it to act so as to expose the receptors of the

organism so that they will acquire certain types of stimuli and avoid others. The stimuli must be acquired in order to avoid prolonged unrest and distress, which may become decidedly malnutritional in their influence, must have the capacity to counter-stimulate the autonomic segment so that it will resume a state of comfortable tonus." (Pages 9-10.) "Emotions and sentiments are cravings that have their origin in the tensions and movements of different autonomic (visceral) segments." (Page 6.) If emotions have a peripheral origin in the characteristic variation of the postural tensions or visceral segments, it is clear that at a given time, the individual, although he may be dominated by a distinct craving, nevertheless is never possessed by a single affective craving. Gradually the autonomic segments become conditioned to seek only such stimuli as are approved by the social group. Will becomes the wish for the esteem of the group. With the integration into a unity of these segmental cravings, together with one's strivings to win social esteem, there develops a personality, the Ego. Normally, of course, only appropriate stimuli would satisfy segmental cravings and so neutralize tension, but unfortunately, it is quite possible for the organism to become conditioned to demand less appropriate stimuli and thereby we have within a single personality the basis for possible conflict between segmental cravings and the striving for social esteem. Out of such conflict arise the mechanisms of suppression, repression, dissociation, and sublimation. In general, then, to Kempf behavior becomes largely a problem of supplying the maximum gratification with the minimum expenditure of energy for segmental cravings, at the same time preventing gratification at the expense of social esteem. Eccentric behavior becomes a compensation for perverse cravings. A segment denied reality must have a substitute. Delusions, hallucinations, and dreams, however bizarre, are but means of gratifying cravings which cannot be gratified by reality.

In appraising Kempf's theory two questions arise: First, what of the validity of the physiological data, the very foundation on which the psychological structure rests? Second, are we justified in considering the affect as purely autonomic in origin? For the physiological data, Kempf finds justification for his point of view largely in the work of Sherrington, Langelaan and Cannon. Even a casual reader might question the sweepingness of certain of Kempf's deductions from the findings of these men. Adequate criticism belongs to the neurologist or physiologist rather than psychologist. So far as the second question is concerned, does not Cannon himself raise the question whether organic changes are sufficiently differentiated in various emotions to account for the affective differences?

As regards Kempf's interpretation of behavior there is indeed much that is illuminating, particularly in his emphasis upon the importance of the Ego's attitude toward the cravings as determining whether his adjustment shall be one of submission, elimination, simulation or sublimation. Certainly such an explanation renders intelligible much that might seem utterly unintelligible, particularly in pathological behavior. Whereas by and large such mechanisms may be acceptable, there are times when the interpretation invites *reductio ad absurdum*. One may grant that Darwin may have had an inferiority complex and yet question whether it be necessary to assume that his use of a high stool in studying was to elevate himself above his fellows, thereby compensating for his feeling of inferiority. One may grant that a patient attempts to compensate for his impotence through the construction of a perpetual motion machine, and yet hesitate to see in invention in

general a compensation for the wastes of masturbation. One familiar with the number of objects on a psychopathic ward which would be available for swallowing, can hardly fail to feel the overemphasis on the swallowing of the ticking watch by the woman patient in simulation of pregnancy. Certainly an interpretive analysis of cases should be substituted for the older description of symptoms. Kempf's "Psychopathology" has been of no little value if it stimulates interest in case analyses. Whether one must ultimately approach all such analyses from the autonomic slant is another question.

As would be expected Kempf makes generous use of the psychoanalytical stock of symbols. In his case studies, their use is what might be expected from any thorough going analyst. However, his interpretation of the symbolic significance of various works of art, suggests that he can see only what he is looking for. For example, he cites Rodin's "Caryatid"—a woman's figure crouched beneath the weight of a burden resting on her shoulder. At once, Kempf sees in the bowl only the psychoanalytic symbol for the uterus, the shoulder is a convenient substitute, and the statue becomes symbolic of the "burdensome uterus and longing for maternity." It is interesting to question to what extent one is justified in reading an interpretation into another's creation. One might also raise the question whether the contemplation of this statue would afford relief from repressions of the type indicated. So far as the psychology of art is concerned, stone obviously from the inherent qualities of the substance itself is adapted to a representation of the type or symbolic rather than the individual or pictorial. But why must all that is symbolic be symbolic of libido at a crude organic level?

Certain it is that no one can contemplate Kempf's interpretations of the mechanisms of human adjustment without finding much that is illuminating and stimulating. Accepted in toto, however, it is legitimate to ask whether such a point of view is conducive to the development of that much to be desired "progressively refining, healthfully constructive attitude toward the fundamental needs and pleasures of human nature." Is an interpretation of life which sees the great driving forces of human behavior as purely organic, conducive to the development of a civilization that will survive? Regardless of its acceptability or rejection on purely emotional grounds, is it not clear that such an ideal would be biologically and ethically a regression? The ultimate test of course is pragmatic. Kempf freely criticizes the "biological castration influence of present American educational and social tendencies" but does not adequately outline a program for their reconstruction. Until he does, it is left to the imagination of the reader to reconstruct a social order in accordance with an ideal of an organic summum bonum. Such an attempt at reconstruction makes particularly clear the social and ethical implications of an autonomic ideal and protects against too uncritical acceptance.

The following books and pamphlets have been received:

- JAMES DREVER.—*The Psychology of Industry*. E. P. Dutton Co., New York, 1921. 148 pp.  
 ———. *The Psychology of Everyday Life*. E. P. Dutton Co., New York, 1920. 164 pp.  
 HENRY R. EVANS. *Educational Boards and Foundations, 1918-20*. Washington Govt. Printing Office, 1921. 11 pp.  
 JULIA WADE ABBOT. *Kindergarten Education, 1918-20*. Washington Govt. Printing Office, 1921. 12 pp.

- Dr. A. A. ROBACK. *Roback Mentality Tests for Superior Adults*. Cambridge, Mass. 14 pp.
- JOHN J. B. MORGAN. *Morgan's Mental Test*. Psychopathic Hospital State University of Iowa. 8 pp.
- HELEN T. WOOLEY, Ph.D. and HORNELL HART. *Feeble-Minded Ex-School Children*. Studies from the Helen S. Trounstone Foundation, 25 East Ninth Street, Cincinnati, Ohio. April 1, 1921. 27 pp. Price 50 cents; 4 or more copies, 25 cents each; in set with 6 other studies \$1.00.
- H. R. BONNER. *Statistics of Universities, Colleges and Professional Schools, 1917-18*. Washington Gov't Printing Office, 1921. 223 pp.
- WM. R. HOOD. *State Laws Relating to Education*. Washington Gov't. Printing Office, 1921. 231 pp.
- CHAS. S. MYERS. *Mind and Work*. C. P. Putnam's Sons, New York and London. The Knickerbocker Press, 1921. 175 pp.
- HARRY DEXTER KITSON, Ph.D. *The Mind of the Buyer*. The Macmillan Co., New York, 1921. 206 p.
- ALEXANDER INGLIS. *Inglis Intelligence Quotient Values*. World Book Co., Yonkers-on-Hudson, New York. 16 pp.





# THE JOURNAL OF APPLIED PSYCHOLOGY

VOL. V

DECEMBER, 1921

No. 4

## AN EXPERIMENTAL STUDY OF CHARACTER TRAITS

By RAYMOND O. FILTER, University of Minnesota

### I

#### STUDY OF SELF-ASSURANCE

##### *Analysis of the Trait*

The looseness of meanings attached to names of character traits demands first consideration. A trait must be defined in order to be studied intelligently. Because definition cannot be made with any degree of accuracy is no good reason for neglect of it. A little attempt to see as a whole and in its various aspects the trait which one studies is sure to be more effective than the procedure of devising tests and naming them as measures of anything at all.

Let the following table I illustrate an attempt at definition. It is a list of typical situations which commonly evoke the reactions that are given this trait-name, together with the reactions comprising the trait. That is, reactions which would represent a high degree of the trait are listed (in the second column) and negative responses, representing a low degree of the trait, are also shown (in the third column). The underlined qualification which appears prominently in the table, OTHER FACTORS BEING EQUAL, demands constant emphasis. Not only self-assurance but a host of other traits and conditions might be responsible for the reactions listed. These other factors must clearly be discounted or equalized before the reactions may be viewed as symptoms or measures of self-assurance. In experiment they must be recognized and evaluated or controlled.

In spite of its incompleteness and its errors (one error will be pointed out later on the basis of experimental findings), the table at once shows its usefulness. One devises tests

TABLE I  
SHOWING TYPICAL SITUATIONS AND REACTIONS INVOLVING SELF-ASSURANCE

Situations	Reactions	
	Other Factors Being Equal	Negative Responses
1. A new situation demanding response.	Undertaking with readiness, carried out beyond demands.	Excessive inquiry—and waiting for directions.
2. A new situation inviting response.	Ready acceptance.	Declined; or needing stronger invitation.
3. Many tasks inviting response.	Acceptance of many.	Carrying a light load.
4. A task demanding preparation.	Tendency to undertake without thoro preparation.	Careful preparation.
5. A task recognized as imposing, difficult.	No hesitation, total engagement, not half-hearted attempts.	Hesitation—partial performance.
6. Responsibility offered.	Accepted.	Avoided
7. Responsibility not offered.	Sought—opportunity made.	Not desired—probably not thought of.
8. Opinion asked.	Readily given.	Modestly withheld or qualified.
9. Estimate of one's own qualities called for.	Given with little delay.	Long pondering.
10. Selection, preferences called for.	Speed of decision.	Slow, wavering decisions.
11. Estimate of one's own capacity to perform certain tasks called for.	Selection of material near limit of capacity, or over-estimation.	Under-estimation.
12. Contradicted when asserting one's own memory of an event, act, etc.	Denial of error.	Acceding.
13. Contradicted in personal convictions by authority and by associates.	Maintenance of conviction (not necessarily public expression).	Suggestible (Re matters of opinion.)

which shall approximate some one or a few of these type situations, and then the table does its further service of preventing generalizations. Successful measurement of some one aspect or element of the trait must clearly be regarded as no measure of the trait. Until its relation to the trait in all of its aspects is determined, no single measure can be safely labelled as a test of self-assurance and be used as such.

### *Experimental Procedure*

In addition to the recognition of the complexity of a character trait there is necessary some insight into its nature as an attribute of individuals. Is it a relatively constant or a variable attribute?

Let us restrict the inquiry at first to just one aspect of the trait, namely to item No. 11 in table I, to expressions of tendencies towards over-estimation. The problem is, to what extent do persons who over-estimate (or under-estimate) their capacities in one situation--on one kind of material--tend to do so in other situations--on other materials?

The different tests which were devised, and applied, vary in value as controlled measures of over-estimation. Only two of them are worthy of complete description. These two tests, however, do seem to possess advantages which should be noted. In the first place, they serve the present purpose very well, of presenting two situations for comparison, which are identical in every respect excepting in the matter of material on which the estimate is based. Test No. 1 calls for estimates of one's capacity to manipulate; test No. 2 calls for estimates of one's capacity to reorganize and reproduce visual impressions; both activities are equally unfamiliar or equally unpracticed (at least in the case of all the subjects to these experiments); and manner of presentation of the material, and all other conditions are identical. In the second place, both tests are objectively scorable.

Test No. 1. STRING FIGURES. (Photographs from Caroline Furness Jayne's book, "String Figures," C. Scribner's Sons, 1906.)

Preliminary: Task was to reproduce with loop of string on fingers patterns shown clear and traceable on pictures. Three patterns, a very simple one and two of representative complexity, were tried, and Subject informed regarding the amount of time he needed for each.

Test Question: Showing S ten more, one at a time, ask: "Which ones of these do you feel sure you would be able to do if allowed 4 minutes for each?"

Criterion of Self-assurance: Number chosen plus number over-estimated (as shown by actual performance tested later).

Measure of Speed of Decision also obtained: Total time from moment No. 1 is handed to S until No. 10 has been chosen or discarded. S not aware of the fact that time is being taken.

Test No. 2. DECOUPAGE. (Binet's test extended.)

Preliminary: Task was to "Draw what the folded and torn paper looks like opened up." Three different patterns tried, the original Binet and two other simple ones. Subject allowed to compare his drawings with the actual pattern.

Test Question: Ten more are folded and torn, and S asked after each one—"Could you do this one correctly, if you were to see it folded once more later, just before drawing it?"

Criterion: Number chosen plus number over-estimated. Speed of Decision measured:—total time of choice—the folding procedure being practically constant for each S.

Dr. June Downey's Will Profile (U. of Wyoming Bulletin No. 3, 1919) contains a test of Assurance, which was utilized in this study. Two sealed envelopes, one stamped M and the other N, are held before the subject for selection. They are then put aside for a while, and when later brought under reconsideration, E makes a point of contradicting the assertions of S relative to which of the two envelopes he had previously chosen. Degree of resistance to contradiction is scored by comparison with standard sample responses. The scoring of this test is not strictly objective as in tests No. 1 and No. 2.

Two tests of Speed of Decision were also given. According to the definition of the trait, table I, items 8, 9 and 10, speed of decision was presumed to be to some extent indicative of self-assurance. Therefore measures of speed of decision were secured in connection with tests No. 1 and No. 2, and two more tests of speed of decision were borrowed. Again Dr. Downey's Will Profile provided a measure, in the Character Checking Test. A list of character traits in pairs is presented and S instructed to check the one trait in each pair which he thinks he possesses. Amount of time employed determines the score.

A modification of Münsterberg's Decision test, the form devised by Sybil Marsh Gibson (*A Decision Study of 150 Young Men and Women. J. of Applied Psych.*, Dec. 1920), constituted the other test of speed of decision. Cards bearing the letters E, I, Y and K, in scattered arrangement, are given S to be examined. S is to decide, without counting, which one of the four letters appears most often on the different cards.

Time and accuracy can be scored, but only time was scored for the purposes of this study.

These tests were given to a group of 26 University students, members of several classes in psychology, and yielded the following results. Disregarding for the present the bearings of speed of decision, and confining attention to the three tests of assurance, the correlations by Rank Difference Method

Between Test No. 1 String Figures and Test No. 2 Decoupage	} $r = .09$
Between Test No. 1 String Figures and Test No. 3 Assurance (Downey)	} $r = -.10$
Between Test No. 2 Decoupage and Test No. 3 Assurance (Downey)	} $r = -.10$

Here is a fair demonstration of variability. If it were established on the basis of a greater number of cases we should conclude that possibly our trait of Self-assurance is purely a function of external factors—of the particular situation—and no constant attribute of individuals at all. What is unequivocally demonstrated is the fact that no one of these tests can be rightfully used alone as a measure of assurance.

Tests No. 1 and No. 2 are of a nature which permits of repetition on the same individuals, after an interval of several months probably, and their reliability may be determined. This has not yet been done.

An attempt has been made to determine the validity of the tests as measures of self-assurance through correlation with estimates. A rating scale was prepared which described six types of individuals, varying in degree of self-confidence, with a grade assigned to each; and then groups of subjects were found who were well acquainted and who provided estimates of each other. The rating scale was successful in serving as a real aid to the judges and in securing very close agreement (M. V. slightly more than a half step) in the different estimates of the same person. But no test results were found to correlate very high with the estimates.

TABLE II		
Correlations with estimates. Two different groups:		
6 boys, members of one fraternity.		
14 girls, members of one sorority.		
Measures of Over-estimation	6 cases	14 cases
Test No. 1 String Figures	$r = .70$	.32
Test No. 2 Decoupage	.70	.28
Measures of Speed of Decision		
Time in Test No. 1	.07	-.14
Time in Test No. 2	.19	-.42
M. Decision Test	-.10	.54
Character Checking	.17	

It became apparent that a mistake was made in attempting to correlate specific test results with estimates on self-assurance considered generally. In order that estimates may serve as a criterion of a test they should be based upon knowledge of activities related to those involved in the test, and upon such activities alone. The scale should be replaced by a number of scales, each descriptive of some different specific kind of self-assurance. In the same manner as the subjects were graded by friends for self-assurance in general (on the scale used), they might be graded for self-assurance in limited spheres of definitely specified types of situations—in matters of athletics alone, or matters of scholarship alone, or in handiwork, or brain work, or in home circles, or in public view, etc., etc.

The following table III summarizes all of the correlations between measures of speed of decision. The correlations are based on as many cases as could be compared, regardless of groups.

TABLE III

No. of cases		
20	Time in String Figures and Decoupage.....	$r$ .38
20	Time in String Figures and M. Decision....	.29
20	Time in Decoupage and M. Decision.....	.28
	(Subject No. 42 was first in speed in all three of these tests.)	
22	Time in String Figures and Char. Checking.	.30
33	Time in Char. Checking and M. Decision...	.15
45	Time in M. Decision and Accuracy in same.	—,50
14	Average rank in Speed of Decision and Estimates .....	—,02

### Conclusions

Evidence showing that the two original tests, No. 1 and No. 2, actually serve as measures of Self-assurance is not at hand. That they possess certain advantages, however, was noticeable to all who studied the tests and observed the reactions to them. The most promising features may be enumerated as follows:

1. Scorable on a strictly objective, quantitative basis.
2. Reactions represent actual attitude of S, not a mere guess without insight.

The trials given are directly comparable with the material estimated.

3. Other factors fairly well controlled.

The objectivity of the material precludes likelihood of much change in reaction under varying subjective conditions.

Test activity uninfluenced by experimenter.

Recognition of purpose of the test makes little difference. Desire to show much confidence in estimating is counter-balanced by desire to show ability to perform the patterns chosen.

Note the correlation of  $-.02$  between average rank in speed of decision and estimates on self-assurance. Though the criterion is faulty, as already indicated, it nevertheless serves for comparison and probably has some value for showing general relationship. Whereas the measures of over-estimation showed low but positive correlations (table II) with the estimates, none of the measures of speed of decision show consistent agreement; and their combined effect (table III) is seen to be entirely unrelated.

A distinction which was not made in Table I, where the trait was defined, is brought to mind by this observation, namely a distinction between decisiveness, or strength of decision, and confidence in one's own powers, or self-assurance in its true meaning. It is a distinction between the manner in which the decision is given, and the kind of decision given. Speed of decision, as a mode of decision, is probably an expression of decisiveness rather than of self-assurance. At least, it does not appear to be the latter. Items 9 and 10 in the table I, defining self-assurance, should evidently be discarded. Whether other measures of decisiveness (such as can be secured by answers to the question, how sure are you? by double checking answers, etc.), whether these show a close relationship to measures of speed of decision, has not been determined. Their relation to measures of self-assurance should also be ascertained.

Quite clearly, too, Speed of Decision seems not to be measured adequately by the Downey Character Checking test, nor by any other single test. Surely when we speak of speed of decision we are as likely to be understood to mean that an individual will decide quickly how much or how difficult labor he can undertake (the consideration under tests No. 1 and No. 2) as to mean that he will quickly decide which trait he possesses (the consideration in the Character checking test). And the results show (table III) that to measure speed in one gives little or no indication of speed in the other type of decision. Such facts should be kept in mind when speed of decision is the subject of conversation.

Possibly if a larger group of persons is brought under examination in a greater number of tests a degree of constancy or generality of the trait may be discovered, but we have no evidence of that possibility beyond the one presented here

in the case of our subject No. 42, who led in speed of decision in three tests, choosing more rapidly than all her comrades in String Figures, Decoupage, and M. Decision test. It is exceedingly regrettable that the Character Checking test had to be eliminated from her group because of mere lack of time. (S No. 42 was ranked 8th, with grade C, for Self-assurance by her friends.)

Speed and Accuracy in the M. Decision test show a distinct negative correlation ( $r = -.50$ ).

The extension of these tests to a greater number of subjects, and also comparison with a greater number of similar tests composed of varied kinds of material, is essential to the definite determination of whether the tendency towards over-estimation is or is not a highly variable and unpredictable trait, and whether or not any one test may be expected ever to serve as a measure.

With reference to the general program of learning something about Self-assurance, it is demonstrated emphatically that we have succeeded no farther than to outline the problem and discover the difficulties that beset its solution. The way for progress is indicated to some extent in our recognition of the need, just expressed above, of completing first the observations relative to the one assumed element of the trait, Over-estimation. We need to know whether or not the tendency to over-estimation can be measured before measures of the total trait can be looked forward to.

The road by which we must climb to insight into character should not be regarded as a highway to be covered with a bound. Quick, quantitative results may establish useful enough measuring devices, but may leave us at the end totally ignorant of what we measured.

### *Summary*

1. Over-estimation, based upon good insight, is a promising criterion of Self-assurance.

2. Tests No. 1 and No. 2 possess advantages for the observation of tendencies to over-estimate. (But the tests are lengthy taking about 20 to 30 minutes each, with the necessary three trials, selection, and then performance.)

3. Tests No. 1 and No. 2 serve excellently for measure of Speed of Decision, on a significant type of decision—"How much can you do?" (Performance not necessary in this case, hence tests can be given in about 10 minutes each.)

4. Speed of Decision is probably not indicative of Self-assurance.



5. Little constancy in speed of decision on different materials is manifested.

6. There is a marked negative relation between Speed and Accuracy in the Münsterberg Decision test.

7. The classification of specific self-assurances and development of separate criteria (rating scales, et al) for each, is a suggested future program which may make possible the determination of validity of tests.

## II

### STUDY OF SPEED OF DECISION

An early study by J. W. Bridges (*An Experimental Study of Decision Types and their Mental Correlates*, Psych. Mon. XVII, No. 1, 1914), concludes with the following assertions regarding speed of decision:

1. Some subjects are very quick and constant in simple situations but are relatively slower and more inconstant when decisions are more difficult.

2. Those quick in subjective decisions are not always quick in objective decisions. (Objective are decisions on matters quite apparent to sense, like judging which card had more holes in it—they would be easily and certainly scored; subjective are decisions on matters of taste and opinion—difficult to score.)

3. In general, practice has no marked effect upon decision time.

In the face of this evidence the use of the two objective measuring devices of character which exist<sup>1</sup> definitely implies a faith in the constancy of character traits, and of speed of decision among them. For when using the scales one gives a grade on speed of decision though only one, or at the most, two kinds of decision have been measured.

Thus the common-sense notion of the trait seems to be carried into scientific usage, where it demands further searching inquiry. In the case of these scales, it may be that the choice of materials has been an exceedingly happy one, and that speed of decisions (in the plural) is fairly well represented in the results (as is evidently assumed); but this seems by no means to have been demonstrated.

Is speed of decision a trait—a general trait, of which we may speak in the singular? Is it a constant and predictable

<sup>1</sup> Reported at the Chicago Meeting of the American Psychological Association, 1920.

(1) A Group Will-temperament Scale, reported by Dr. June Downey.

(2) Test No. IX, Form 1800, of the Bureau of Personnel Research of the Carnegie Institute of Technology, reported by M. J. Ream.

characteristic of an individual's reactions, regardless of the nature of the decisions called for? Are there individuals who show themselves to be quick in decisions of all kinds? Or of many kinds? Or of customary kinds? Are there other individuals who excel in one or a few kinds of decision, while they prove ordinary or very slow in other kinds?

Such, and related, questions must be answered before the hope of securing any single representative measure of speed of decision can be entertained.

It was seen that Bridges has given a negative reply to the main questions suggested above. His answer, however, may be regarded as not decisive, inasmuch as it was arrived at by observation of only 19 subjects. It gives only a suggestion of the probable nature of the trait in the general population.

However, the investigation on Self-assurance just reported in the foregoing section, disclosed again a condition of lack of constancy in the speed of decision of a group of 20 persons, and more in some cases. None of the time correlations is high, .38 being the highest found between any two tests. But with this observation and its support of the view in favor of a general lack of constancy in speed of decision, the very noticeable fact appears that one subject in the group of twenty took first rank in speed in as many as three different tests. Such a fact discourages the hasty generalization regarding the inconstant and unpredictable nature of the trait. Possibly larger groups than those tested so far contain more of such individuals who distinguish themselves by the possession of general superiority, or inferiority, in speed of decision. The present investigation is now an attempt to determine the strength of this possibility by varying and extending the observations.

The aim is to measure the reactions of a large group of individuals to a variety of decision-problems, and to observe whether constancy or variation in speed is the rule; to observe also if possibly the general rule for the mass hides any significant groups of exceptions.

#### *Description of the Tests*

A brief statement of the instructions for each test, and where possible an illustration of the material, is given here. In practice the instructions were read aloud to the subjects, and the tests were bound together in the form of a pamphlet. No test was seen before signal to start was given. After one test was completed all subjects listened while instructions for the next one were read. So through all six.

# AN EXPERIMENTAL STUDY OF CHARACTER TRAITS 307

Test I. Check the one trait in each pair which you possess:

Patient	Impatient
Cautious	Reckless
Dull	Clever
etc.	

Test II. Write in small square below each group the most frequent letter:

This test has been described in the foregoing section, as the Munsterberg Decision Test modified by Sybil Marsh Gibson.

Test III. Grade these actions as follows:

If VERY VERY GOOD.....	VVG
If VERY GOOD.....	VG
If GOOD.....	G
If INDIFFERENT.....	I
If BAD.....	B
If VERY BAD.....	VB
If VERY VERY BAD.....	VVB

To be an active church member.....

To praise your own cooking.....

To tip your barber regularly.....  
etc.

Test IV. Check the one item in each pair which in your opinion has the greater value *to the human race*:

tea kettle.....flatiron.....

science.....music.....

mirrors.....keys.....  
etc.

Test V. Compare each of the figures below with the small square above the line. Write an S, L, or E inside of each figure—

S if smaller than the square in area  
L if larger than the square in area  
E if equal to the square in area

Test VI. Which is the greater evil, generally? Check one of each pair. Do not skip, but take one pair at a time in order:

To use cosmetics (paint)..... To chew gum noticeably in public.....

To frequent burlesque..... To frequent bull fights.....

To hunt out of season..... To be instrumental in re-instating capital punishment.....  
etc.

Test I is the Downey test for speed of decision, with a few additional traits added to the list. Test II is the M. Decision test, modified and transformed from a card sorting test to a paper test, but retaining its essential characteristics. Tests III, IV, V, and VI are original contributions, though clearly patterned after many of the questions and problems that have been devised to test moral disposition and appreciation. One feels a distinct freedom from the hazardous interpretations of the student of moral disposition in employing material of this kind in a manner which disregards the nature of the responses given. For here the concern is only speed, and no interest is attached to the values of the replies. Of course, the subjects taking the tests are instructed in a manner to produce careful consideration on their part. Even purely hypothetical norms of performance are referred to as actual in the instructions in an attempt to make the subjects feel the need of careful judgment, and to make accuracy of performance appear the significant consideration in the score.

The different tests call for a variety of decisions. They are planned to bring into play a number of different kinds of factors—to stimulate a variety of the typical considerations—that tend to make people hesitate when deciding. Test I is distinctive in its demand upon self-reference, and it invites all the difficulties of introspection and the hesitations due to modesty in assertion. Test II presents objective, easily verifiable material, but it deprives the subject of the method by which he would ordinarily determine his answer. (He is not permitted to count the letters.) The question differentiating persons in this test seems to be: "Can I trust the hasty impression?" Test V involves decisions of the same objective kind, and the same causes for hesitation.

Tests III, IV, and VI are of a kind in presenting subjective material for judgment—that is, the decisions are to be based on matters of opinion, on social and moral considerations of a high degree of complexity, on personal convictions. Test III differs from test VI only in its demand that grades be assigned as answers, in place of the mere check on a paired comparison. It calls for absolute judgments in contrast to the relative judgments of test VI. Test IV has an advantage among these tests of so-called subjective decisions in reducing the element of speed of reading which possibly affects the scores on these tests.

*Results*

A time limitation was set on the tests, and amount of work—number of decisions—scored.

The number of records obtained was 142. They represent the reactions of a very homogeneous group of university students, mostly under the Junior year in school, though including a few advanced students. The records of the latter, however, show no noticeable deviation from the rest.

The time limitations for the various tests were determined by preliminary trial upon a group of 25 students, and set below the fastest performances in that group, the aim being to have no persons finish the test. Unfortunately this trial group, an evening class in Applied Psychology, did not contain the most speedy performers, and as a result the tests were launched out upon the regular class groups with a time limitation too lenient. Quite a number of individuals finished some of the tests. This circumstance causes a distortion of the distributions of scores which must be referred to later when results are interpreted.

INTERCORRELATIONS OF TEST RESULTS. 142 CASES. PEARSON PRODUCT MOMENTS FORMULA

	Test I	Test II	Test III	Test IV	Test V	Test VI
Test I.....	...	.43	.47	.56	.31	.51
Test II.....	.43	...	.51	.50	.66	.32
Test III.....	.47	.51	...	.55	.45	.40
Test IV.....	.56	.50	.55	...	.57	.46
Test V.....	.31	.66	.45	.57	...	.37
Test VI.....	.51	.32	.40	.46	.37	...

These coefficients of correlation are not as useful as are other summarizations of the results which follow. In the first place the distortion of distributions—the above-mentioned massing of scores at the top of each scale, due to the fact that the time allowed for the tests was too great—has an influence upon these correlations which is difficult to estimate. It must lower it somewhat, though probably not a great deal. In the second place, this influence is not equally strong upon each of the correlations in the table, and comparisons must be accordingly uncertain.

If the coefficients are accepted at their face value, or with some allowance for the attenuating circumstance, they would seem to indicate a degree of constancy in speed of decision. Those persons who are quick in decisions of one kind tend to be quick also in decisions of quite another kind. The correlations are not high enough, however, to warrant the con-

clusion that any one test would serve as a measure of speed of decision. This is the important conclusion which may be safely drawn from these figures. It seems almost absurdly simple as the objective of elaborate investigation, but after all, the demonstration serves a good purpose. It is essential to the success of character testing that pictures of this kind be presented, for without them exalted opinions regarding the worth of single tests are apt to prevail. Later figures will show how unfortunate would be a classification of persons according to score on one or two tests of this trait.

The two tests which show the highest correlation with one another are numbers II and V, the two which have been described as objective and as being in that respect similar to each other and quite different from the others. Evidently this feature is reflected to some extent in the higher correlation. It means that a number of the individuals at least show a relatively greater constancy in speed in decisions of an objective nature than in the other types of decision. However, another similarity of the two tests, besides their objectivity, should be noted, for it may account for the higher correlation. The two tests are, after all, visual perception tests, and it is likely that equality of visual endowment, or more likely still, equality of practice in gauging by visual impression, may to a large extent determine the correlation.

The table of intercorrelations shows a still further distinction between these two objective tests and the others. If the averages of the columns be taken, with the rows II and V omitted, the figures resulting are as follows:

Test I	Test II	Test III	Test IV	Test V	Test VI
.513	.440	.473	.523	.425	.456

These figures show the averages of the correlations of each test with all others excepting II and V, and it becomes clear that tests II and V have the lowest average correlation with the rest of the tests. On the whole there is less correlation between the speed in objective types of decision and the subjective types than there is between the different subjective types here tested. This corroborates one of the observations of Bridges, reported at the beginning of this section.

A further very simple statistical treatment was applied to the test results, which yielded returns of considerable value for insight. It supplied a means of classifying the individuals tested, it presented a more intimate picture of the actual meaning of the correlations (uninfluenced by the disturbing attenuat-

ing circumstance), and it allowed a further comparison of the tests. The procedure was as follows: On the basis of the various distributions the grades in the various tests which represented most closely the 20 percentile score and the 80 percentile score were determined. These served then to differentiate the upper and the lower 20% individuals from each other and from the middle (60%) group. The choice of the 20 percentile as the distinguishing level was made after a scrutiny of the various distributions of scores. It was found by trial that the 80 percentile fell upon an even score in most of the tests, that is, definite groups were more regularly divided at about this point than at other points (like the 25 percentile) which were tried. The lower group is really marked off by the 23 percentile score. Thus the individuals were classified as being either speedy, slow, or ordinary in each test. On the score sheet, then, it became a simple matter of drawing a circle in black ink about all scores above the 80 percentile score, and a circle in red ink about those below the 20 percentile score. A small section of the score sheet is reproduced here to illustrate the effect of this procedure.

Subject	SCORES					
	Test I	Test II	Test III	Test IV	Test V	Test VI
No. 125.....	5	7	22	25	20	12
No. 126.....	9	13	13	29	22	13
No. 127.....	25	24	39	48	34	20
No. 128.....	16	5	25	38	33	10
No. 129.....	10	24	21	37	23	7
No. 130.....	7	18	34	39	10	18
etc.						

Now to count the number of circles of different color which characterize the different individuals becomes an easy matter. The following table summarizes the counts. The first column shows the number of black circles, that is, the number of tests in which the individuals were among the upper 20% for speed; the second column shows the number of individuals who attained that status (who had that number of black circles to their credit). The third column and the fourth column, respectively show the same facts for the lower 20% (number of red circles) as columns one and two.

No. of tests	No. of persons high	No. of tests	No. of persons low
6	6	6	2
5	2	5	6
4	6	4	10
3	10	3	16
2	21	2	20
1	30	1	27
0	66	0	58

0 high and 0 low (middle grade in all tests)  
(with no circles of either kind).....12

This table indicates that 6 persons were in the upper 20% of the group of 142 in all 6 tests (like No. 127 in the section of the score sheet on the preceding page), 2 persons were in that group in 5 of the tests, 6 in 4 tests, 10 in 3 tests, etc.; and among the slow individuals, 2 were exceedingly slow in all 6 tests, 6 in 5 tests, 10 in 4 tests, etc. There were 12 individuals who failed to reach either the fast or slow status in a single one of the tests.

Clearly, there are here discovered two groups of individuals, and probably really three, to whom the general attribute of speed of decision, in some one of its forms as quick, slow, or mediocre, may be referred without much danger of inaccuracy. The one group of 6 individuals definitely possesses the quality of quickness in decision, as far as this number of tests gives evidence of it. Another group of 2 individuals may be labelled slow; and a third group of 12 individuals are characterized by constant mediocrity of speed of decision. These distinctly recognisable groups comprise 4%, 1%, and 8%, respectively, of the total number of students tested.

The degree to which the positions of the less pronounced individuals were interchanged from fast to slow and to middle groups in the various tests can also be determined readily by observation of the circles on the score sheet. A significant question which such a determination answers, for instance, is the following: Did the two persons who fell below the upper 20% in only one test fall way down into the slow group in that test? The same question is answered at the same time with respect to those who were high in all but two tests, all but three tests, etc.; and also the places taken by those who were low in all but one test, all but two tests, etc., can be determined. The answers are tabulated in the following manner. The table really constitutes an analysis of the correlations. It shows the fluctuations in position of individuals at the various grades or levels of possession of the trait. The



first line, if written out completely, would read as follows: 100% (2 individuals) of the 2 who were high in 5 tests (in all but one test) were mediocre in that one test.

TABLE SHOWING FLUCTUATIONS IN POSITION

## High Groups.

100 %	high in 5 tests are mediocre in 1 test.
100 %	high in 4 tests are mediocre in 2 tests.
80 %	high in 3 tests are mediocre in 3 tests.
10 %	high in 3 tests are mediocre in 2 tests; low in 1 test.
10 %	high in 3 tests are mediocre in 1 test; low in 2 tests.
57 %	high in 2 tests are mediocre in 4 tests.
29 %	high in 2 tests are mediocre in 3 tests; low in 1 test.
14 %	high in 2 tests are mediocre in 2 tests; low in 2 tests.
43.3 %	high in 1 test are mediocre in 5 tests.
40 %	high in 1 test are mediocre in 4 tests; low in 1 test.
13.3 %	high in 1 test are mediocre in 3 tests; low in 2 tests.
3.3 %	high in 1 test are mediocre in 2 tests; low in 3 tests.

## Low groups.

100%	low in 5 tests are mediocre in 1 test.
100%	low in 4 tests are mediocre in 2 tests.
94%	low in 3 tests are mediocre in 3 tests.
6%	low in 3 tests are mediocre in 2 tests; high in 1 test.
60%	low in 2 tests are mediocre in 4 tests.
20%	low in 2 tests are mediocre in 3 tests; high in 1 test.
15%	low in 2 tests are mediocre in 2 tests; high in 2 tests.
5%	low in 2 tests are mediocre in 1 test; high in 3 tests.
30%	low in 1 test are mediocre in 5 tests.
44%	low in 1 test are mediocre in 4 tests; high in 1 test.
22%	low in 1 test are mediocre in 3 tests; high in 2 tests.
4%	low in 1 test are mediocre in 2 tests; high in 3 tests.

This table repeats facts, but it serves to make a few more assertions possible. It is really not only the meagre 4% (6 individuals) at the top of the scale, and the 1% at the bottom of the scale, who show constancy of speed in decisions. Inasmuch as the table shows 100% clear from slow decisions in both the group of 2 individuals who were high in 5 tests and the group of 6 individuals who were high in 4 tests—in the first two rows of the table—it might be regarded as no great

distortion of fact to include the latter 8 individuals with the 6 of the upper group who are characterized by general constancy of speed. This raises the percentage of individuals who may be called truly speedy in decision to about 10%.

On the same considerations the slow group may be extended to include two more groups who showed 100% clear from fast decisions, and the percentage of truly slow individuals may be raised from 1% to about 13%. Further extensions than these would certainly not be justified by the facts shown in the table.

### CONCLUSIONS

Correlations between test results are positive and fairly high, indicating that individuals who are quick in decisions of one kind tend to be quick also in decisions of other kinds. This does not constitute a demonstration of high degree of constancy of speed of decision, however, as later qualifications show. There is little likelihood that any one or two tests can be developed to measure the trait adequately. Whether or not this small battery of six tests may serve as a fair measure is not ascertained, and it remains an inviting inquiry.

In corroboration of Bridges' observation, to the effect that those persons quick in subjective decisions are not always quick in objective decisions, it is found that the correlations between speed in the two objective decisions and all others are lower than those between any of the other subjective decisions and the rest. Most constancy of speed (the highest correlation) is found between the two objective tests.

Percentile classification reveals three groups of individuals with whom a certain degree of speed of decision seems to be constant. The groups are very small, however. By lenient extension of classifying boundaries, the group which may be regarded as constantly speedy in decision (according to these tests) comprises 10% of the total number of students tested. 8% of the total number are constantly mediocre, and about 13% show constancy in slow decision. Strictly speaking, only 4% appear dependably fast, 8% mediocre, and 4% regularly slow.

A general criticism of the group testing situation should be added as an interpretation of the results obtained. The experience with individual tests of this same trait, with the tests on Self-assurance reported in the preliminary study, served to make the group testing situation appear unsatisfactory in several respects. The group test introduces an uncontrollable and undesirable factor of competition (not usual in the deciding-situations of life), and it reduces all

tests to a level of similarity to some extent, thus encouraging high correlations.

These two factors are probably more influential in determining these correlations than would appear to the observer not acquainted with the finer advantages of the individual test situation. With regard to the factor of competition, very clearly the rush to beat one's neighbor, or, disregarding the social influence as a disputable item, the rush against time, which is undeniably operative, makes of the test a measure of something more complicated than its name implies. One's decisions under stress of that kind are doubtfully like those (in speed) which one ordinarily makes. The instructions, the admonitions regarding the need of careful judgments, which are given to offset the adjustment favoring speed, only complicate matters. To the less suggestible and the more competitive belong the spoils.

The particular advantage of the individual test in this respect is the possibility of presenting the material in a manner, so that the subject may be kept ignorant of a time element in the test. Unless the essential feature of natural time-rate of decisions is kept in the situation a test cannot very well be said to measure natural time-rate of decisions. Yet that, it seems, is what we mean when we speak of speed of decision.

The other disturbing condition in the group testing situation is not an inherent obstacle to the success of group tests of speed of decision as the other factor is, but it is possible that its influence operated to some extent on the results of this investigation. For here was an attempt to present under a single cover, and in rapid succession, at the same sitting, and under the very same conditions of mood and motion, sets of problems for decision, whose principal value was supposed to lie in their variation from each other. At least, the correlations of the responses were to determine whether decisions of a variety of kinds tended to be elicited at relatively constant or variable speed from the same individuals. Surely, if the situations were identical in some such fundamental respect, any characteristic of the separate tests, or of the deciding activities responding to the tests, which might favor or incline toward variability, would be much reduced in effect. This possible influence is mentioned only as a reminder of the very doubtful character of the correlations found. They are not high as they stand, but if this similarity of conditions had been avoided in some way, so that the true nature of varied decisions had gained expression, it is possible that a greater degree of inconstancy would have been demonstrated.

## SUMMARY

1. Fair positive correlations show some constancy of speed of decision.

2. The difference between objective and subjective decisions is reflected in the correlations. Those persons quick in one type are less apt to be quick in the other type than in other decisions of the same type.

3. The large majority (fully 69%) of individuals tested cannot be classified as either quick, slow, or mediocre. They are too inconstant. Only 10% may be characterized as quick, 8% as mediocre, and 13% as slow, according to these tests.

4. No single test differentiates these relatively constant individuals from those whose scores fluctuate from extreme to extreme in the various tests.

5. Group tests have inherent disadvantages for measuring this trait, or for improving theoretical insight into its nature.

## CAPITULATION

1. Logical analysis of Self-assurance reveals Over-estimation as an element which seems to be particularly amenable to experimental test.

2. Two tests are reported, which were tried and found to possess certain advantages for observation of tendencies to over-estimate. They also provide measures of natural speed of decision.

3. The trait of self-assurance, according to evidence of its nature gained by application of a big variety of tests, is still too imperfectly defined to enable the adequate testing of the tests by definite check against objective criteria. (The criteria are not determined.)

4. Development of a servicable criterion in the form of accumulated estimates, to aid which a rating scale was supplied, was successful in the matter of securing reliable estimates, but complexities of the trait, again, deprive the estimates of value as a criterion. Hence

5. The classification of specific forms of self-assurance and development of tests for each separately, is suggested as an improvement upon the attempt to deal with the trait in general. As criteria, then, either objective and directly measurable expressions of the trait may be discovered, or more useful estimates will be obtainable—estimates which shall definitely bear upon the features which the test seeks to measure, and upon these alone.

6. Speed of decision is probably not a reliable indicator of self-assurance.

7. Inconstancy in speed of decision is demonstrated to be the characteristic of the majority of students tested. Constancy seems clearly to be the exception rather than the rule, even under conditions favoring constancy.

8. Speed of decision is not likely to be measured by any one test or by a few tests, excepting in individuals of proven constancy.

9. Group tests have inherent disadvantages for measuring speed of decision, or for improving theoretical insight into the nature of the trait.

# PSYCHOLOGICAL NORMS AMONG UNIVERSITY FRESHMEN

By ANNA KELLMAN WHITCHURCH

This investigation was made at the psychological laboratory of Northwestern University during the year 1916-1917. It was undertaken for the purpose of finding whatever correlation may exist

(1) Between the results in the performance of certain psychological tests, and the mental ability of Freshman students as shown in the attainment of grades in Mathematics;

(2) Between the grades in Mathematics and the ability of the Freshmen as shown by the quarter of the High School class from which they came;

(3) Between the grades in Mathematics of the Freshmen and their instructor's independent judgment of their intelligence.

## *Description of Tests Used in the Present Investigation*

The tests we used were (1) Easy Directions Test,<sup>1</sup> (2) Hard Directions Test,<sup>2</sup> (3) African Proverb Test, (4) Substitution Test.

We used the easy directions test first, and its sections were pasted together in such a way that forty continuous directions were given. The directions were not numbered, however, as they are in the plate.

The hard directions test followed that. This was given exactly as Woodworth and Wells used it.

The African proverb test was arranged in the laboratory of Northwestern University. The object in view was to develop a system of simple tests by which to gain an insight, if possible, into students' capacities for grasping differences and similarities in abstract problems.

The substitution test has been used often in one form or another. The form devised by W. F. Dearborn and described by Whipple, "Manual of Mental and Physical Tests," Part II, p. 136.

---

<sup>1</sup> Woodworth and Wells Tests No. 88071B, Stoellting.

<sup>2</sup> Woodworth and Wells Tests No. 88071, Stoellting.

*Subjects and Method of Investigation*

The Department of Mathematics allowed one hundred Freshmen who were taking Mathematics AI to devote the time of one class period to the business of taking the tests. They, together with a class of sixty Sophomores and Juniors, met in a large auditorium. The chairs had desk arms, and each student was provided with a lead pencil in order that all conditions of work might be as uniform as possible. Four professors from the Department of Psychology, three from the Department of Mathematics, and seven members of a Psychology Seminar assisted in giving these tests.

All the test blanks were marked on the back with a check at the top. Those who passed the blanks passed them face down but with the check mark in the correct place. As soon as the papers were received by the subjects they wrote their names and addresses on the back. The papers were left face down until all instructions were given. Then the subjects were told to grasp the paper at the bottom in such a way that it could be conveniently turned. At a given signal the papers were turned, and at the word "begin," the tests were started.

When the time for a test was up, the completed papers were passed along the rows of seats to the right. The papers for the new test were passed in the same direction, so that a minimum amount of time was lost in the mechanics of getting the papers collected and passed. This made possible the giving of the four tests mentioned above and the following tests by another experimenter, in slightly more than an hour's time: two opposites tests, two completion tests, and an information test.

With so many helpers the amount of unfair work among the subjects was reduced to a small per cent. On the whole, we felt that the results were as good as could be gained from the ordinary group test. For the most part the subjects entered into the tests seriously and tried to make them worth while.

The directions for each test were given orally as follows:

1. *Easy Directions Test*

"On the other side of this paper you will find forty simple directions to follow. They are like these, I have on the black board.

"a. Add one dot to the largest group.

"b. Write any number smaller than 16.

"You are to work as rapidly as possible; for you will have only a minute and a half in which to do this test. The test is so simple that you will not be likely to make mistakes. However, work for accuracy as well as speed. Are there any questions? Start as soon as you are told to begin, and stop immediately when you are told to stop. Ready—turn papers—begin." The stop watch was started at the word "begin."

### 2. *Hard Directions Test*

"This test is like the one you have just finished, save that the directions are harder to follow. There are fewer of them, and you will have a minute and a half in which to work, as you did before. You will be more likely to make mistakes here; so pay more attention to accuracy, but work as fast as possible. Ready—turn papers—begin."

### 3. *Proverb Test*

The following diagram was put on the board:

<hr/>	
<hr/>	
<hr/>	
<hr/>	
English	African
{ }.....	1.....
{ }.....	2.....
{ }.....	3.....

"You will find on this paper two columns of proverbs arranged as in this diagram—a column of English proverbs on the left, and a column of African proverbs on the right. The African proverbs are numbered; the English proverbs have empty parentheses beside them. You are to turn your paper, read the directions at the top and follow them. Work carefully; for one mistake will count much against you. Are there any questions? Ready—turn papers—begin."

Three and a half minutes were allowed for this. The results showed that that time was too long; for more than half finished the test. This fact will be discussed later.

### 4. *Substitution Test*

"At the top of this paper you will find nine circles, each containing a number and a symbol. (This was illustrated.) Below, you will find two columns of figures of five places with five empty squares arranged horizontally beside them. You are to write in the squares the symbols corresponding to the numbers as given in the circles above. Finish each number before you start the next. Finish the first column before you start the second. You will have four minutes in which to do this. Every thirty seconds I shall say 'Mark,'



and you are to put a diagonal mark at the exact place where you are then working. This will enable us to find your improvement from time to time. Are there any questions? Ready—turn papers—begin."

### *Method of Scoring Tests*

In the Easy Directions Test there are forty directions to be followed. The record of speed was made by counting the number of directions followed. The accuracy was figured on the basis of the number done.

$$\frac{\text{Number of errors}}{\text{Number done}} = \text{Accuracy}$$

The Hard Directions Test had approximately twenty-one directions, and it was scored in the same way as that used for the Easy Directions Test.

The Proverb Test was scored for speed by counting the number done, and accuracy was found by counting off  $12\frac{1}{2}$  for each proverb placed incorrectly.

In order to secure the results in speed in the Substitution Test, one was counted for every square filled. A record was made of what each person accomplished in the eight periods of thirty seconds into which the time for the test was divided. So few errors occurred that they were negligible. The improvement which each person made was found by subtracting what was accomplished in the first two periods from what was accomplished in the last two.

The records for all the tests were then placed on cards, one card for each individual and his records. An example is given on the next page.

The records for each subject in all the tests were ranked, and the rankings placed in red ink above the first records. This was done that the Spearman formula might be used in correlating the tests. The rankings of each person in each test were then added. These results were ranked and formed the general intelligence record. Placing the records on cards in that way, facilitated the ranking to a great degree, because each record was easily moved from place to place. All the data, too, for each person was eventually on one card. This made the records easy to use and easy to refer to.

### *Specific Errors*

We shall now discuss the errors made in response to the four tests respectively. Many of the errors in the Directions Tests were due, of course, to hasty reading.

Freshman		
Smith, Mary		
	Speed	Accuracy
Easy D.	40	5
Hard D.	21	24
Proverb	7	14
	Improvement	
Substitution	147	12
Combined Ranks	405.5	

Sample of card used in recording results in tests. Speed = number of directions finished, proverbs placed, and squares filled. Accuracy = per cent of error in completed work. Improvement = number of spaces filled in last two periods minus number filled in first two. Combined Ranks = rankings in all the tests added together.

In the Easy Directions Test there were surprisingly few errors. Thirty-nine people made mistakes, and altogether there were fifty-seven errors made. Twenty-six people made one error each; ten made two each; two made three each; and one made five errors. In speed these people ranked well in comparison with the others tested. Those having one error ranked from  $5\frac{1}{2}$  to 96. The former had completed twenty directions and the latter thirty-seven. Those who made two errors ranked in speed from 11 to 100. The former had completed twenty-two directions; the latter had finished all forty. Those having three errors ranked from  $29\frac{1}{2}$  to 38. The one person who made five errors ranked  $79\frac{1}{2}$  in speed.

The errors taken in order are as follows:

In No. 3, one person placed the number 9 instead of 2 in answer to the question: "How many ears has a cat?"

In No. 5, instead of a cross, two people merely put a line through the word *hat*.

In No. 6, one person wrote out the word eight, and another put figure 8 at the thick part of the line.

In No. 7, two people wrote a word of one letter.

In No. 8, four people put a cross in one of the squares instead of a dot; one put a dot in all the squares; one put a cross in one, and a dot in another.

In No. 9, one person crossed out all the words.

In No. 11, one person drew a curved line like a hill; another connected those given so that a hill was formed.

In No. 12, six people said there is one *t* in *twist*.

In No. 13, three people crossed out the undotted line; one put a dot in the center; and another dotted and crossed it.

In No. 14, seven people wrote *o* above instead of after the larger number.

In No. 18, one person crossed out the middle square.

In No. 19, one person put a larger number instead of a smaller one after 10.

In No. 21, one person underlined the *g*.

In No. 22, one person wrote *z* above the dots; one put a dot between.

In No. 23, two people said 12 feet make a yard.

In No. 25, one person put a dot above the line.

In No. 26, two people multiplied instead of adding the numbers; one simply rewrote them.

In No. 27, one person rewrote the whole word *John*.

In No. 28, four people dotted the circle.

In No. 29, one person wrote *F*, and another wrote *C*.

In No. 32, one person crossed out the egg-shaped figure.

In No. 33, two people made only one dot.

In No. 35, two wrote *N* instead of *e*.

Some of these errors are alike and may be accounted for. The errors in No. 5, where the lines were drawn instead of crosses made, were probably due to the fact that the direction just preceding called for a line to be drawn. Then, in No. 8, where five people put a cross in the white square, the error was probably made because a cross had been asked for just previously in No. 5, while no dots as yet had been called for. The largest number of errors made in any one direction occurred in No. 14, where seven people put the *O* above instead of after 86. This was probably because in the preceding direction they were asked to put a dot above the line. In No. 28, four people put a dot instead of a cross in the circle. This undoubtedly was due to the fact that in No. 22 and No. 25 dots had been called for. In No. 32, one person put a cross in the egg shaped figure, probably because No. 31 had called for a cross.

These errors seem to show the influence of the preceding actions. The memory seems to have been carried over and to have inhibited the new reaction.

In the Hard Directions Test sixty-two people made errors; and one hundred and twenty errors were made altogether.

Thirty-four people made one error each; ten made two; nine made three; seven made four; one made five; and one made six. The rankings in speed of those making one error ranged from 4 to  $89\frac{1}{2}$ . Those making two errors ranged from 4 to  $89\frac{1}{2}$ . Those having four ranked from 7 to  $89\frac{1}{2}$ . The one having five ranked  $89\frac{1}{2}$ , and the one having six had that ranking also.

The errors in each direction were as follows:

In No. 1, two errors were made. One person drew a line above one letter, and another dotted all of them.

In No. 3, two people said Christmas comes in March.

In No. 4, fourteen answered correctly, but in answer to No. 5 ten of them crossed out their answers, showing that the next direction was incorrectly followed. One wrote nothing, and one said the sun rose in the west.

In No. 6, fourteen said a horse has two feet; one left the space blank.

In No. 7, three wrote, "NO."

In No. 10, thirteen left this blank, and one wrote, "Yes."

In No. 8, three gave the correct number of days.

In No. 11, four left this blank.

In No. 13, five made a square, and one made a cross.

In No. 14, one person made two crosses, and one made one.

In No. 15, one wrote 6; one wrote 1000; one wrote 1076; two left it blank.

Six people left No. 16 blank.

Eight said that nights in summer were longer than in winter.

In No. 19, one answered incorrectly.

Two wrote 12 in No. 20.

In No. 21, sixteen people wrote the letters in the wrong place or else wrote the wrong letters.

The mistakes in this test were due clearly to misreading and confusion. There is nothing about the errors made to show that the effects of one process are carried over to the next.

In the Proverb Test, forty-one people made errors, and eighty-four errors were made. Twelve people made one error each; nineteen made two; six made three; and four made four.

Those who made one error ranked from  $3\frac{1}{2}$  to 62 in speed. Those who made two mistakes ranked from 1 to 62. Those who made three errors ranked from  $19\frac{1}{2}$  to 62; while those making four also ranked from  $19\frac{1}{2}$  to 62.

In No. 1, no one made an error.

In No. 2, seventeen errors were made. No. 5 of the African Proverbs was given for its equivalent in meaning six times; No. 2 was given eight times; No. 7 once, and No. 3 twice.

In No. 3, four errors were made. No. 2 of the African Proverbs was given for its equivalent three times; No. 3 was given once.

In No. 4, sixteen errors were made. No. 4 of the African Proverbs was placed here seven times; No. 6 five times; No. 1 twice; No. 3 once.

In No. 5, fifteen errors were made. No. 3 of the African Proverbs was given as its equivalent nine times; No. 5 three times; No. 6 once; No. 4 once; No. 2 once.

In No. 6, seven errors were made. No. 6 of the African Proverbs was given as its equivalent twice; No. 3 twice; No. 7 twice; No. 4 once.

In No. 7, there were eleven errors. No. 1 of the African Proverbs was given as its equivalent four times; No. 3 four times; No. 5 once; No. 2 once, and No. 6 once.

In No. 8, fourteen errors were made. No. 7 of the African Proverbs was given as its equivalent eleven times; No. 2 once; No. 5 once; No. 4 once.

No. 3 of the African Proverbs was misplaced nineteen times. No. 7 was misplaced fourteen times.

In the Substitution Test few errors were made. Three people did not mark the thirty-second periods as the signals were given. One person worked across the page instead of down the first column and then down the next.

The Spearman Foot-Rule Formula was used to obtain the correlations. The formula is

$$R = 1 - \frac{6 \sum G}{N^2 - 1}$$

*R* stands for index of Correlation.

$\sum G$  stands for the sum of the numerical gain in rank of the individuals in the seconds, as compared with the first series of the results which were correlated.

*N* stands for the number of individuals in the series.

The 4 was obtained by the use of Thorndike's table, which converts *R* into the Product-Moment equivalent *r*. Thorndike's formula for so converting *R* is

$$r = 2 \cos \frac{\pi}{3} (1-R) - 1$$

Following are the correlations between the results of the various tests:

Easy Directions and Hard Directions (Speed).....	70
"                    "                    "                    (Accuracy).....	33
Easy Directions and Proverb                    (Speed).....	42
"                    "                    "                    (Accuracy).....	33
"                    "                    "                    Substitution                    (Speed).....	52
Easy Directions (Accuracy) and Substitution (Improvement).....	36
Hard Directions and Proverb (Speed).....	36
"                    "                    "                    (Accuracy).....	36
"                    "                    "                    Substitution (Speed).....	36
Hard Directions (Accuracy) and Substitution (Improvement).....	00
Proverb and Substitution (Speed).....	32
Proverb (Speed) and Substitution (Improvement).....	25
General Intelligence and Easy Directions (Speed).....	71
"                    "                    "                    (Accuracy).....	32
"                    "                    "                    Hard                    "                    (Speed).....	72
"                    "                    "                    "                    "                    (Accuracy).....	32
"                    "                    "                    Proverb (Speed).....	54
"                    "                    "                    "                    (Accuracy).....	55
"                    "                    "                    Substitution (Speed).....	48
"                    "                    "                    "                    (Improvement).....	44

### *Estimated Abilities*

We obtained an estimate of the natural ability of each student from the Professors in Mathematics who had under their instruction the Freshmen who had been tested. The following instructions were given to them:

So far as you are acquainted with the following persons, mark down after each name a number indicating the intellectual fitness of each individual. Grade in terms of what you consider their real ability or capacity to do difficult intellectual work—not in terms of their customary class performances.

Mark	+ 2	after superior individuals
"	+ 1	" very good "
"	0	" merely average "
"	- 1	" just below "
"	= 2	" very poor "

In that way an estimate of the ability of all but four of the hundred was obtained. Each professor was asked to make as nearly as possible a distribution corresponding to the following: Out of ten pupils, four would rank at zero, two at +1, one at +2, two at -1, and one at -2. The professors' estimates are shown in Table I, pp. 328-329.

From the high schools where these students had graduated, we learned whether these people had been in the first, second, third or lowest quarter of their class at the time of graduation. We obtained the desired information from all but seventeen high schools. The other ranks are given in the third column of Table I.

*Arrangement of Results in Table*

The group of one hundred Freshmen was divided into four quarters. The results of the tests were arranged in order of attainment when all the ranks of each student in the tests were added together. These are given in Column 1, Table I. Thus the first quarter has the twenty-five who did the best work in the tests; the next quarter has those who came next to them; the third and fourth quarters are arranged accordingly. The next column, No. 2, contains the estimates of the professors upon the abilities of the students in question. In the next column, No. 3 are their respective positions in the high school class with which they graduated. In the last column, No. 4, are the grades obtained in Mathematics. These grades were given in terms of A, B, C, D, F,—A being the highest possible grade, and F denoting failure in the subject. For convenience's sake, these grades are given in the table as 1, 2, 3, 4, 5,—1=A, 2=B, etc.

The following correlations were made between General Intelligence as found in the combined results of the tests, the grades in Mathematics, the estimates of ability by the professors, and the quarter of the high school class in which they were placed:

General Intelligence and Mathematics' Grade	=	.24
" " " Professor's Estimate	=	.22
" " " High School Quarter	=	.36
Mathematics Grade " " "	=	.55
" " " Professor's Estimate	=	.92
Professor's Estimate " High School Quarter	=	.59

The following table is divided into four quarters according to the respective rankings in the tests. The second column gives the professor's estimates; the third gives the High School Quarter from which the students came; the fourth gives the Mathematics grades.

TABLE I  
FIRST QUARTER

	Sum of Ranks In Tests	Prof. Est. Terms of +2, +1, 0, -1, -2	H. S. Quarter	Math. Grades
1	614	+2	1	1
2	597	+1	1	1
3	572	+1	1	2
4	567	+1	3	4
5	558	+2	1	1
6	557	+1	1	2
7	533.5	-1	1	4
8	528	+2	1	1
9	526	+1	1	2
10	525.5	0	4	3
11	525	+2	1	1
12	523.5	+1	1	2
13	521.5	+1	.	2
14	520	+1	2	3
15	517.5	+1	.	2
16	514	-1	1	2
17	511	-2	4	5
18	510.5	+1	1	3
19	504.5	0	.	3
20	503.5	-2	2	5
21	503.5	0	1	3
22	496	-1	3	5
23	496	-1	.	4
24	493	.	.	.
25	488	+1	1	3

SECOND QUARTER

	Sum of Ranks In Tests	Prof. Est. Terms of +2, +1, 0, -1, -2	H. S. Quarter	Math. Grades
26	483.5	+1	1	.
27	483.5	.	.	2
28	480.5	0	2	.
29	480	+1	4	2
30	477.5	0	2	1
31	472	0	1	4
32	467.5	0	1	2
33	462	-1	1	3
34	461.5	0	.	4
35	461.5	-2	2	3
36	461	0	1	5
37	458.5	+2	4	1
38	453	-1	2	1
39	450.5	+1	4	5
40	440.5	0	1	2
41	440	0	1	3
42	435.5	0	3	2
43	427	-2	.	5
44	425	+1	2	5
45	423.5	-2	1	2
46	422	0	1	4
47	421.5	-2	2	3
48	415.5	0	3	4
49	414.5	+1	1	2
50	412.5	-1	4	2
		0	1	5
				3



PSYCHOLOGICAL NORMS AMONG UNIVERSITY FRESHMEN 329

TABLE I—Con.

THIRD QUARTER

	Sum of Ranks In Tests	Prof. Est. Terms of +2, +1, 0, -1, -2	H. S. Quarter	Math. Grades
51	412.5	0	1	3
52	408.5	0	2	3
53	403	+1	1	2
54	403	0	4	3
55	401	-1	1	3
56	396.5	0	3	3
57	395	-1	.	4
58	395	0	1	3
59	394.5	-1	2	4
60	388	+2	1	1
61	380.5	+2	1	1
62	377	0	.	3
63	374.5	+1	1	3
64	368	0	2	3
65	367.5	-2	4	5
66	367	+1	1	2
67	365.5	+1	2	2
68	364	+1	2	3
69	356.5	0	2	3
70	345	...	.	.
71	340	-1	3	4
72	338.5	+1	1	2
73	332	-2	4	3
74	332	+1	1	2
75	331	0	.	3

FOURTH QUARTER

	Sum of Ranks In Tests	Prof. Est. Terms of +2, +1, 0, -1, -2	H. S. Quarter	Math. Grades
76	324.5	+1	1	2
77	324.5	+2	1	1
78	318.5	-1	3	5
79	318.5	+1	2	2
80	313	+1	2	1
81	311	0	1	3
82	310	+1	.	3
83	303	-1	1	4
84	293.4	+1	1	3
85	291	+1	4	2
86	284	-1	3	5
87	283	...	.	.
88	275	+2	1	1
89	269.5	+1	1	1
90	262.5	-1	.	5
91	245.5	-2	1	5
92	240.5	-2	.	5
93	238	0	3	3
94	238	-1	1	5
95	203.5	0	1	3
96	201.5	0	1	3
97	196.5	+1	2	2
98	184.5	+1	3	2
99	161	-2	4	5
100	144.5	-2	4	.

In order to see, if possible, the difference in mental ability between those who ranked highest and those who ranked lowest in the tests, it was decided that three from the top and three from the bottom should be tested individually. Results for only two at the bottom were obtained. The tests used for this purpose were: Healy Puzzle Box; The Dutch Homestead (test for memory of ideas from Whipple's Manual); a Part-Whole and a Genus-Species test.

The next table shows the results obtained. The numbers in the first column correspond to those in the first column of the quartiles in Table I.

The column entitled "Steps" under the Healy Puzzle Box Test shows the number of steps necessary to open the box. A perfect performance is six. Under "Time," the number of minutes and seconds required are shown.

The columns devoted to the Dutch Homestead show the number of words and ideas recalled. In the test there are one hundred and eighty words and ninety-four ideas. The median reaction time is given for the Part-Whole and Genus-Species tests.

TABLE No. II

	Healy Puzzle Box		Dutch Homestead		Part-Whole Median Reaction	Genus-Species Median Reaction
	Steps	Time	Words	Ideas	Time= Seconds	Time= Seconds
1.	23	5' 57"	159	47	2.2	2.3
2.	10	2' 31"	94	25	2.6	1.6
3.	17	3' 50"	67	27	2.7	2.2
98.	8	1' 54"	135	32	3.3	4.25
100.	17	4' 42"	106	27	2.0	2.5

The results from the above do not correspond with the results of the first set given. For instance, the man who was number 98 in the first series used the fewest steps and least time in opening the Puzzle Box. As the least possible number of steps is six, his performance was almost perfect.

In the Dutch Homestead test the girls who ranked first in the other tests were superior. The other four are nearly alike in the number of ideas remembered, while in the number of words given the two lowest are distinctly better.

The reaction times in the Part-Whole test vary but little. In the Genus-Species test, practically the only variation is between Number 2 and Number 98.

The result hoped for was not obtained. These last four tests show little if any difference between those who rank high in the first tests and those who rank near the bottom

### *Introspective Data from Tests*

There has been, up to this time, too little discussion of the mental capacities or functions that are most active when one is reacting to the tests in which we are interested. There is too little description, too, of methods by which different reactors go about their work. This data can be obtained in part by introspection. This is important because before a test can be evaluated it is necessary to obtain all the information possible about it.

With this in mind, one day around the seminar table, eight people, six advanced students and two professors in psychology, took the tests we had given to the Freshmen. Afterward each one recorded his mental processes while taking the test. These people were experienced in introspection, and it was hoped that we might find some of the frequent ways in which those people tested ordinarily went about their work, and the various mental processes used in doing the tests.

The following are some quotations showing various interesting ways of proceeding with the test material. The first are from the Easy Directions Test.

"I had a feeling of pushing from within or of an effort to get ahead in a short space of time. There was also a feeling of checking myself lest I should omit a line or make another error. My attention was directed rather narrowly to the directions themselves."

The method just quoted is one probably used by many people. The feeling of push would be prevalent in most tests where speed was involved. The next person went about the test in a different way.

"I read rapidly. I did not pronounce all the words, and I am not sure that I even saw them all. I was conscious of trying to understand the direction, and a feeling of surety and satisfaction came with the grasp of meaning. After the meaning came, I waited with gaze fixated on the symbols at the end, and the proper movement of the pencil resulted automatically. When the feeling of meaning failed to come, I consciously repeated parts of the direction or even the whole thing over again. There was consciousness of strain throughout."

The two people quoted seem to represent two groups often seen in Mathematics classes. The one is careful and method-

ical and sure. The other breathlessly grasps an indefinite meaning and automatically works a problem by the first method that comes to mind. He works more rapidly than the first, and though he may make more mistakes, will accomplish as much in the end because of his speed.

In regard to mental processes, the following statements were made:

"The test seems to call for readiness in recalling alternatives. The presence or absence of hesitancy would soon be felt."

"The requirements seemed simple enough until No. 7 was reached. This called for a word of three letters, and I felt like checking myself to see which word might be selected. In No. 11, where we were asked to mark the line which looks most like a hill, I wondered that any of them appeared like a hill. My next notion was that the turning of the paper might make one of them into a hill. I finally chose the third as the one being most appropriate when the paper was in its natural position."

"The figures on the right received the most passive attention except in the cases of No. 6 and No. 11, where I had to discriminate between stimuli. In No. 11 I was puzzled. In no other case did I *consciously* discriminate. It seems to me that the predominant function here is voluntary attention to immediate sensory stimuli."

"No. 10 called for voluntary inhibition. Nos. 1, 2, 4, 6, 8, 11, 13, 16 and 18 called for appreciation of objects. Nos. 5, 6, 14, 15, 19, called for comparison of values. Nos. 3, 7, and 12 called in association. No. 9 required general memory."

All of the above quotations seem to show the required use of attention and apperception. Probably more than any other one thing, this Easy Directions Test does test the function of attention.

Some remarks in regard to the Hard Directions Test follow:

"The words of the instructions continued to sound auditorily-kinaesthetically while the answers were being written. They seemed to determine the proper answers. Occasionally I started to make the wrong answers. Immediately an uncomfortable feeling arose. Pertinent parts of the instructions became more vivid auditorily. With the right answer came a feeling of relief."

"Here I had to keep attention not only on what was passing under my eye, but usually upon what had preceded; and not only that, but I was always in an attitude of expectance also—i. e., toward the future. Consequently I was embarrassed

always by alternating attention between the directions, and I was all the more fearful of making mistakes. There was a feeling of greater effort here than in the Easy Directions Test, where I attended only to what was passing under my eye."

"This test required perpetual attention. A perception organized into a very definite appreciation or apperception of what was wanted in each case was necessary. This was different from the easy test and noticeably more difficult in that a constant resistance to suggestion was required throughout. Imagination seemed to play no part here. Memory was not so much involved since the facts called for were of a triflingly common nature. But power to command this knowledge instantly and without forewarning and to write it as quickly as it was thought was needed."

"Reasoning was most used in this test, as many times reasoning compelled me to do the opposite from what memory suggested as being correct for certain of the questions. All the time I felt that I would have to be careful lest the simplicity of the questions as asked would lead me to erroneous answers. The feeling of caution was present."

"There was present inhibition to respond in the usual way. This sometimes in the first reading led to closer attention. Often I tended to the wrong answer, but stopped to compare my knowledge with the suggestion made on the paper."

At once we wonder what the effect of the various reactions called forth in individual subjects would have upon the speed and accuracy of their performance. In the Easy Directions Test the close attention which was mentioned several times would surely lead to accuracy, and would probably steady the rate of work. In the case of the man who waited automatically for the answer to come to his pencil, one might suppose the process of answering to have been subconsciously done. While this might accelerate speed, it would be apt to block accuracy.

In the Hard Directions Test, the subject who held his attention closely to the direction upon which he was working, to the one preceding that, while consciously alert for the one about to come, probably made no mistakes, and his speed would be greater than the one who read carelessly and hurriedly and so had to go back. The fact that answers were called for contrary to one's usual line of thought would decrease speed and lay one liable to errors. The constant alertness and caution necessary to overcome that would make for accuracy, but would probably decrease speed. In short, any-

thing purely automatic could be very rapidly accomplished. Those directions calling for caution, reason and discrimination would cause delay.

To proceed to the Proverb Test:

"My consciousness of the directions seemed to be summed up in a visual kinaesthetic feeling of putting numbers in brackets. I read through an English proverb without getting a clear comprehension of its meaning—i. e., I did not translate it into other terms or apply it mentally to any concrete situation. Nevertheless, I proceeded immediately to glance hastily through the African proverbs. I rushed up and down the list in great haste without attempting to read one of them through, and without system, until I struck a phrase or turn of words that 'felt right.' Then I automatically put the number in the brackets. The consciousness of agreement between the meanings was a vague process."

Another wrote, "The first time I undertook this test I was embarrassed by the Directions. The nature of the test had not been adequately explained to me. I wanted to make speed, but I had to look into the proverbs to understand the directions. Then the truth flashed over me, that I was adding time to my record. I was embarrassed, and because of that I started under a handicap. When I was ready to go at the work I followed this method: I fixed the first English Proverb in mind, and then read through the African proverbs comparing each with the meaning of the English proverb. After selecting one fitting its meaning, I proceeded the same way in each other case. In each case the proverb suggested a visual image, and I found myself looking for a common or near common element in these images and comparing them. When once I was ready to write the desired numbers beside the English proverbs I had the feeling of having solved a problem. The ability called for here seems to be that of handling imagery more or less abstract."

One student used the method described above, but instead of reading an English proverb first, he selected an African proverb and searched for one of like meaning in the English column. After describing his method he says: "But with many of the African proverbs I felt that the meanings were too ambiguous ever to admit of close comparison. Continued thinking and reasoning led me to believe that I had stressed the details too much, and that I should select that larger meaning of each proverb and adhere to that for comparative purposes."

Another says: "The process is one of *comparison* analagous to that of *superposition* in geometry. Superposition here is one of ideas rather than symbols. Selection of the correct parallel proverb is a matter of assembling all the African ones not yet used as possibilities. Next, the irrelevant ones must be rejected. In carrying on the process of selection, the one proverb has to be borne in mind as the problem or end in view."

Here again we have two types of individuals, the one who hastened to read here and there in the two columns, and the one who conscientiously and carefully read and compared one proverb with the others. In the final results one type did as well as the other. It would seem from the descriptions we received that much the same process was carried on as that used in solving a problem in mathematics. It was necessary first to analyze a proverb and retain its abstract meaning. Comparison of one abstract meaning with another was necessary all the way through. The final selection would adequately compare to the solution of a problem. The mind acting accurately in this situation would doubtless do equally well with a mathematics problem.

Some of the remarks on the Substitution Test follow:

"The feeling that I should need to remember the symbols later led me to try to impress them upon my mind. It was soon found to be economical to try to memorize the easier symbols. An association helped in fixing the symbols. For instance, I thought of 7 as  $\triangle$  incomplete and up-side-down."

Another says he made no conscious effort to remember. He simply relieved memory of that strain and allowed a habit to form. He had to inhibit the desire to place one symbol for another that had just been used.

A unique way of doing the test is described here: "The first time I tried this test I did not play the rules of the game, but read the number, held the digits in my mind's eye and then hunted for the necessary symbols in succession. These symbols I was soon able to write in the squares without looking at the squares. No attempt was made to learn the symbols, and they were not learned at all well. I had their location in mind much as a typist has in mind the location of the keys. Primary memory for the numbers seems to be the only mental function I employed. Later I found the real learning of the symbols more difficult. Primary memory for the whole "key board" was finally acquired, largely by aid of imaginary association in which I imagined houses, trees, etc., in a yard standing for the symbols respectively."

The way this man did the test the first time may have been duplicated by members of the Freshman group who were tested. In that case accuracy would be perfect, and speed would range high. The test would not be one of learning at all. It would test rather primary memory and motor facility.

The next person made it a real learning test. He says: "I kept my eyes on the key as much as possible and filled in the spaces without looking. I read each individual digit separately and then looked at the key. At first it was necessary to hunt consciously for digits in the key. Gradually the proper digit with its symbol would tend to jump out from the key as a whole; i. e., I would tend to turn directly to the proper place in the key. Occasionally when I got lost I hunted up and down the key board without taking in much of it. As learning progressed, the symbols for the different numbers became consciously recognized and identified when found. The ability to find them quickly seemed to be correlated with this fact that the symbols were getting learned."

One student said: "The power to form visual images rapidly and accurately seemed to come into play. After writing two numbers, it became easier to 'See' the proper symbol mentally as soon as the digit was read than to look back to the heading. To people whose memory images are other than visual, I believe this would be a difficult test, and could be done only slowly."

As in this case, visual imagery would assist greatly in the test, but motor imagery might be quite as adequate.

From this introspective data we find the predominant mental functions used in these tests to be: active attention, response to immediate sensory stimuli, discrimination, comparison and selection, and association. Looking back at my own experience in mathematics, I would say that these functions played a large part in solving problems. It is necessary to center one's attention to analyze and to discriminate in order to come to a solution.

Professor Frank Nugent Freeman in his "Psychology of the Common Branches" tells us that problems are solved by analyzing conditions and following up clues. Before the solution can be obtained it is necessary to get an understanding of the problem in concrete terms. In geometry, for instance, a pupil must thoroughly understand that the lines and angles represent real spacial relations.

This analysis and following up of clues is apparent in the Directions Tests and the Proverb Test. So far as we can



rely upon the introspective data, it would seem that these tests tested the same ability as that which Professor Freeman feels is required by the student in Mathematics.

### *Discussion of Results*

The data gained from those who took the tests for the purpose of introspection seem to show that the tests really tested some of the abilities necessary for successful work in Mathematics. However, the correlation between the results of the tests and the Mathematics grades is low (.24). This may be due to poor adjustment of the students to their work during the first semester of their University course, or it may be that the introspections did not reveal the real abilities tested. If the latter is true, then the tests selected may not test the abilities necessary in work in Mathematics.

On the other hand, it is quite possible that the students tested did not put all of their energy into the tests, and therefore did not do them as well as they might have done. It is impossible to tell just how much attention the subjects did give their work. While they appeared to work hard during each test, they undoubtedly did not put in their maximum attention. Indeed, it is impossible to measure the absolute degree of attention by any laboratory methods at present perfected. This inability to measure attention is a drawback in every phase of laboratory psychology. Granting that the students did not do their best in these tests, while the tests might have got at mathematical ability, the results would not show high correlation.

Three Professors graded the students upon their work in Mathematics. It is unlikely that all three of these have the same standards of grading. Therefore, the grades themselves with which these correlations were made are more or less arbitrary.

Then, again, the tests given individually to those who stand highest and those who stand lowest do not show a very great difference. It is well known that high correlations are not to be expected from a homogeneous group of subjects. Either part do well in those things in which others fail, or all fail and succeed in the same lines, and thus balance the results. This causes the results of the group to fall into a narrow distribution. Our group of subjects were all Freshmen. This pre-supposes a degree of homogeneity. The subjects have all passed the same entrance requirements to begin with. By referring to Table I it will be seen that forty-five of the one hundred are from the first quarter of the high school class

with which they graduated. Granting that some of these classes were small, and that the quarter in which each student was placed was assigned more or less arbitrarily by the principals of the high schools, still the fact remains that forty-five per cent of these students were considered considerably better than the average in their classes. This fact of the homogeneity of the group would account for a relatively low degree of correlation.

The high correlation (.92) between the grades in Mathematics and the professor's estimates of the students' abilities were to be expected. The grades received by the students probably influenced the professors when they gave the estimate, and vice versa.

The correlation between the grades in Mathematics and the high school quarters from which the students came was .55. This was as high as could be expected when we remember that forty-five per cent came from the first quarter.

On the whole, this investigation seems to have been worth while in spite of the fact that the results in the mental tests did not correlate highly with the Mathematics grades. The student who did the best work in the tests was found to be an exceptional student. She received A grades in all her work for the first semester. The student who stood lowest in the first tests failed in all his work. He failed twice in Mathematics, and, after having been tutored privately, failed the third time. Even with the unavoidable difficulties that come up in getting accurate results from mental tests, it would be possible for a teacher to gain some insight into his students' abilities by the use of such tests at the beginning of a semester's work together. This investigation shows a more or less definite connection between the class standing and abilities involved in doing the tests.

Dr. Wissler<sup>3</sup> who made correlations with some of the testing done at Columbia University, felt that the markings in college classes would correlate with mental tests to only a slight degree. Miss Calfee's<sup>4</sup> correlations between the tests she gave at the University of Texas and the average class standing were low. Professor Bell<sup>5</sup> in his work at the same University found his highest correlation between a class standing and the results of a test to be only .31. Miss Rowland and Miss Lowden<sup>6</sup> found a correlation of .25 between grades

<sup>3</sup> *Psych. Rev. Mono. Sup.* 3, 1901, No. 6.

<sup>4</sup> *J. of Ed. Psych.* 4, 1913, 223-231.

<sup>5</sup> *J. of Ed. Psych.* 7, 1916, p. 381.

<sup>6</sup> *Jour. of Exper. Psych.*, Vol. 1, No. 3.

in Mathematics and one group of tests. While the results recorded here are much like those we found with our work with Northwestern University Freshmen, still we feel that there is a hopeful outlook. With perfection of methods of giving mental tests, and perfection of a system of grading for class work, the connection between the two will be more evident. When that time comes, the results of mental tests will be an index to the abilities of students in their class work.

### *Summary*

1. The highest correlation in this investigation was .92. This was between the professor's estimates of the students' ability and their grades in Mathematics.

2. The lowest correlation was .0. This occurred between the accuracy in the Hard Directions Test and the improvement in the Substitution Test.

3. Among the tests the correlations ranged from .72 between General Intelligence and Hard Directions Test to .0 between the accuracy in the Hard Directions Test and the improvement in the Substitution Test.

4. The correlations with the Mathematics grades were:

Mathematics and General Intelligence.....	.24
Mathematics and High School Quarter....	.55
Mathematics and Professors' Estimates...	.92

## A PRELIMINARY REPORT OF AN EXPERIMENTAL ANALYSIS OF CAUSES OF STUTTERING<sup>1</sup>

By LEWIS O. ANDERSON, University of Wisconsin

Of human afflictions, stuttering is one of the most elusive. Persons who are handicapped by it often find no one who can give intelligent advice or treatment. Many become victims of quacks. Moreover, if we look to various recognized authorities on stuttering we find great diversity in the theories which have been put forward. Repressed "complexes" have been held responsible. Weak imagery has been assumed. The trouble has, in general, been looked upon as mental, but there has also been belief in some defect of physiological coordination. Inherited neurotic dispositions, emotional instability, abnormal fears, trauma, imitation, and many other things which have been present in various cases, are alleged to be causal. Indeed, stuttering has appeared in such an unusual variety of conditions and circumstances, and often with so many complications, that it is not at all surprising to find disagreement as to what the causal factors are. There is a host of scattered hypotheses. Usually several factors are considered operative at the same time. But which, and how many? And which of these are dominant? Definite knowledge is lacking. There have been many scattered observations, but there seems to have been no systematic survey made on the basis of experiments set up to put hypotheses to a test and thus get evidence on a scientific basis.

One important outcome of the numerous observations is a growth of the view that stuttering is not a disease in itself

---

<sup>1</sup> This investigation has been in progress for two years, being under the authority of the Speech Clinic, University of Wisconsin, of which Doctor Smiley Blanton is the Director. The experiments (at the University of Wisconsin) were performed in a special research room in the psychology laboratory. Professor Joseph Jastrow and Professor Clark L. Hull are to be thanked for their great interest and valuable suggestions. The author is especially indebted to Professor Hull for careful criticisms during the experimental work and in the preparation of the manuscript. The subjects, with speech disorders, for these experiments were obtained from the following sources: (1) cases at the Speech Clinic, University of Wisconsin, (2) cases at the Speech Clinic, United States Public Health Service Hospital, Waukesha, Wisconsin, and (3) cases at the Wisconsin Industrial School for Boys, Waukesha, Wisconsin.

but only a symptom. If, then, it is not at its root essentially a speech defect, investigation of it must not be limited to vocal reactions. Mental conditions apart from speech must be studied. The purpose of this article is to present briefly some results of a number of special tests designed to detect and measure certain traits in reactions outside of the field of speech, and given in such a way as to show, if possible, whether these traits are definitely associated with stuttering. In view of the large number of alleged causes, this experimental work was planned so as to constitute a survey of significant types of mental and physiological processes which are common to vocal and non-vocal reactions. While there are, no doubt, both innate and external contributing factors, these tests were aimed principally at those which are innate, or at least internal. An essential part of the method consisted in a detailed comparison of groups of stutterers with groups of "normals." Fortunately, it was also possible to include a group of ex-stutterers. At the University of Wisconsin, practically no treatment preceded the tests in these subjects.

The tests whose results are reported here are briefly described as follows:

1—Foot tapping:

If there is a general defect of physiological coordination, it could reasonably be expected to manifest itself in a tapping test, especially if the variability of tapping rate be measured. A foot tapping test was therefore set up in such a way as to get a graphic record of the separate taps.

2—Hand coordination:

It is well known that stutterers talk well under imposed rhythm. A test was then designed for coordination and rhythm in other than speech reactions. The subject went through certain complicated series of arm movements, in some trials following the beating of a metronome, and in other trials simply going as fast as he could.

3—Block test:

In view of alleged weak imagery, this and the two following tests were introduced. This one was devised as a test of memory span for movements. Four small wooden blocks were used.

4—Two-dot test:

This test was aimed to detect weakness in the kinaesthetic field. With eyes closed, the subject made a certain arm movement (making a dot with a pencil on a piece of paper), and about one second later tried to reproduce the movement exactly. He was scored on accuracy of reproduction, judgment

of his error, and on constant error (i. e. tendency to err with the second movement constantly in a given direction).

#### 5—Visual imagery:

A few subjects were given visual imagery tests, some of which, e. g. the Binet Simon clock-test, are well known.

#### 6—Complex reaction-time:

A familiar characteristic of the stutterer's reactions, at least in speech and emotional situations, is instability. He is variable. He goes ahead by spurts. To possibly measure this instability, a typewriter with a device to expose the digits 1, 2, 3, and 4, one at a time in irregular order, was set up in such a way as to get a graphic record of sixty reaction-times in uninterrupted succession.

#### 7—Inhibition:

The characteristic spasms and blocking on words clearly suggest some defect of inhibition. A test was accordingly devised to measure inhibition outside of vocal reactions. A list of numbers in irregular order was read, and the subject was directed to respond by raising his arm slightly every time "3" was read immediately after an odd number but not after an even number. Forbidden responses, and also partially inhibited forbidden responses, were recorded.

Tables I and III summarize typical results of the subjects at the University of Wisconsin. Since certain changes were made in the tests before they were given to the second of these two groups of subjects, the results in the two cases are not always strictly comparable. Table II summarizes results from the boys at the Industrial School at Wakesha. It should be stated that the hand-coordination test and the inhibition test were slightly simplified for the subjects at the Industrial School and at the U. S. Public Health Service Hospital. These results will now be discussed in the order in which the tests appear in the Tables.

An inspection of the results of the first test in Tables I and III shows no consistent difference between the stutterers and the normals in tapping rate. Ex-stutterers, however, are very rapid (Table III). A significant difference<sup>2</sup> exists be-

<sup>2</sup> In order for the difference between two averages to be significant, the difference should be at least 2.5 times its probable error. This probable error is obtained from the probable errors of the two averages by the formula:—

$$P. E. D = \sqrt{P. E. ^2_{M_1} + P. E. ^2_{M_2}}$$

Throughout this article the term "significant difference" thus means one which is at least 2.5 times its P.E. In the above mentioned case it is 3.2 times its P.E.

tween their average and that of the normals. This type of reaction, then, seems to be definitely associated with recovery from this speech defect. As to variability of tapping rate, the improved stutterers have a much lower score than the normals. Obviously, there is no evidence here for any defect of coordination at this level, that is, of the rapid but coarse muscular movements of tapping.

The next test, that of hand coordination, shows a characteristic difference between stutterers and normals. A careful comparison of Table II with Table III will make clear that the stutterers tend to do relatively better when following the beating of the metronome, while the normals tend to do relatively better without the metronome. "No. right" means number of trials passed without error. The results from ten stutterers at the hospital show the same tendency, but taken alone they are not sufficiently reliable for a conclusion. At any rate, it must be noted that the tendency, as far as it appears in these results, is in clear harmony with the well known effect of imposed rhythm on the stutterer's speech. The significance of this is that the trait appears to be general: here it seems to play an important rôle in reactions which have practically nothing to do with speech. If we compare the improved with the unimproved stutterers, another significant difference is found. It would appear from these subjects that the stage of improvement in speech essentially involves a marked disturbance in the types of coordination required in this test. This, however, needs further investigation. As to individuals, one patient who was troubled by being speechless from fear in military situations, did very well in this test and remarked that he had very poor control in e. g. driving a car in tight places. He accounted for his doing well in this test by saying that there was no fear connected with it.

The block test (Table III) shows the stutterers to be distinctly inferior to the normals in memory span for movements. The scores in this test were also averaged separately for severe and mild cases; these averages were respectively 24.1 and 30.4, with a significant difference. The question arises as to what bearing this has on theories of weak imagery. To consider this it is first necessary to ask if this is a test of mere imagery as such or whether it is a test of other functions along with imagery. Memory span tests have been used as tests of attention. What this test really tests is probably the ability to hold several things in mind at the same time. The stutterer here appears to have difficulty in grasping several

TABLE I

SHOWING RESULTS OF THE FIRST GROUP OF SUBJECTS AT THE UNIVERSITY OF WISCONSIN

Test	Type of Record	NORMALS			STUTTERERS		
		Cases	Score	P. E.	Cases	Score	P. E.
FOOT TAPPING	Number of taps	18	132.4	3.94	12	121.2	3.69
	Variability (seconds).....	18	.19	.011	12	.153	.016
HAND COOR- DINATION....	No right	Without metronome	.76	....	17	.82	....
		With metronome	1.8	....	17	1.7	....
COMPLEX RE- ACTION TIME	Number of errors	17	3.9	....	16	4.9	....
	Time (seconds).....	17	.79	....	16	.79	....
	Variability (seconds).....	17	.55	.049	16	.456	.029
INHIBITION....	Forbidden responses.....	15	3.1	....	9	4.6	....

TABLE II

SHOWING RESULTS OF THE SUBJECTS AT THE WISCONSIN INDUSTRIAL SCHOOL FOR BOYS

Test	Type of Record	NORMALS			STUTTERERS		
		Cases	Score	P. E.	Cases	Score	P. E.
HAND COOR- DINATION....	No right	Without metronome	.83	....	21	.38	.094
		With metronome	.83	....	21	.76	.14
BLOCK TEST...	Number right..	13	34.7	1.41	21	27.1	.7
TWO-DOT TEST	Motor accuracy (mm).....	11	12.7	.68	20	11.6	.32
	Error of judgment (mm).....	11	14.45	1.46	20	15.45	.71
	Constant error (mm).....	11	8.8	.83	20	6.2	.4
INHIBITION....	Forbidden re- sponses.....	11	6.55	.89	15	7.93	.59
	Partially inhibited	11	3.65	.48	15	1.8	.22



TABLE III  
SHOWING RESULTS OF THE SECOND GROUP OF SUBJECTS AT THE UNIVERSITY OF WISCONSIN

Test	Type of Record	NORMALS			STUTTERERS						EX-STUTTERERS		
		Cases	Score	P. E.	Unimproved			Improved			Cases	Score	P. E.
					Cases	Score	P. E.	Cases	Score	P. E.			
FOOT TAPPING	Number of taps...	39	121.7	2.3	11	128.6	....	10	128.9	....	11	140.3	5.3
	Variability (seconds).....	39	.206	.009	11	.192	....	10	.167	.009	11	.186	....
HAND COOR- DINATION...	Without metronome	27	1.11	.085	8	1.37	....	7	.57	....	12	.92	....
	With metronome	27	.63	.098	8	1.37	....	7	.71	....	12	1.08	....
COMPLEX RE- ACTION TIME	Number of errors.	24	3.87	.368	8	2.12	....	6	2.	....	11	2.73	....
	Time (seconds).....	24	.891	.019	8	.797	.025	6	.879	....	11	.778	.025
INHIBITION....	Variability (seconds).....	24	.6041	.031	8	.52	.062	6	.5637	.049	11	.3754	.017
	Forbidden responses.....	36	3.11	.25	12	2.9	....	9	2.11	.41	13	2.54	....
	Partly inhibited.....	36	4.69	.3	12	3.25	.41	9	4.44	.58	13	3.69	.51

movements simultaneously. The significance of the kinaesthetic factor seems to be that in consciously directed speech movements a great many kinaesthetic elements must be held in mind simultaneously and thus coordinated. This the stutterer is apparently unable to do. To the writer the most logical conclusion seems to be that the trouble lies in the coordination of imagery rather than in mere imagery as such; that the stutterer, especially in certain situations, lacks the ability to coordinate the kinaesthetic and other imaginal elements involved in speech movements when they appear in large numbers simultaneously.

In the two-dot test the only significant difference found between the stutterers and the normals (Table II) was in "constant error," the stutterers having the lowest score. Apparently, they have fully as good control as the normals of these movements made with the eyes closed. There does not, then, seem to be any weakness in the kinaesthetic field, that is, as far as arm movements are concerned.

The results of the visual imagery tests given to twelve stutterers at the Hospital showed more correspondence with intelligence than with severity of stuttering. Of the two men who did much better than the others in his test, one was a college graduate and the other had attended a university. They were both very bad stutterers. One other stutterer had practically recovered; he succeeded in only one out of the seventeen points scored in this test. Those who appeared lowest in intelligence did poorly in this test. Now it can be said of this test too that it tests "attention" along with imagery. But it differs radically from the above-mentioned "block test;" the instructions were repeated, if necessary, so that it would not really be a test of "span," that is, ability to hold a large number of things in the first grasp.

In the complex reaction-time test the stutterers do not seem to differ uniformly from the normals either in speed or in number of errors (false reactions). As to variability of reaction-time, the results do not show any strong indication that stutterers differ from normals. This test accordingly fails to give any evidence for greater instability or irregularity of reactions in stutterers than in normals. But there are, no doubt, characteristics of instability in other types of situations not reached by this test. Ex-stutterers, again, differ markedly from all the other subjects. They are unusually regular, as shown by their low score in variability of reaction-time. Here, then, is a trait which seems quite definitely associated with recovery from this speech defect.

## EXPERIMENTAL ANALYSIS OF CAUSES OF STUTTERING 347

The inhibition test shows no tendency in the stutterers to differ uniformly from the normals in mere amount of inhibition. Among certain stutterers, however, there was observed to be some individual correspondence between lack of inhibition in this test and lack of inhibition in speech and other behavior. According to teachers at the Industrial School, some of the boys who showed much lack of inhibition in this test showed the same characteristic lack of inhibition in their behavior on the school grounds. One, e. g., was quick to retaliate when out playing with the other boys. He had a "don't care" attitude, though he was much improved; he freely told, in a good natured way, how stubborn he had been when formerly brought into court. Another boy who showed much lack of inhibition in this test was flighty and constantly getting into trouble; and stuttered only in rapid speech, not in reading. A boy who showed less lack of inhibition than the average in this test, differed also in the form of his stutter: he hesitated and showed signs of hidden struggle in getting each word out. A stutterer, of another group, who showed much lack of inhibition in this test was troubled mainly by a strong tendency in all his behavior to react out of proportion to the intensity of the stimulus. Many of those who showed least lack of inhibition in this test (i. e. had fewest forbidden responses) showed great inhibition in their speech movements.

Partially inhibited responses are seen (Tables II and III) to be more numerous in the normals than in the stutterers. This relative scarcity of partial inhibitions in the latter shows itself clearly in both groups where recorded, while the number of completed forbidden responses mentioned above varied in different groups of stutterers. This indicates that the stutterers in these groups differ as to whether they will start or not start a wrong response, but they tend to agree in their inability to check or modify the response after it has started. It seems that the stutterer's response tends to be of the all-or-none type; that the impulse does not get started at all as overt response or, if started, it cannot easily be stopped midway or modified. It must run its entire course. In this connection it must be remembered that in speech there is an immense number of impulses at play. If, then, there is decreased ability to check or modify impulses midway in their course of expression, it is reasonable to expect greater disturbance from conflicting impulses and to expect blocking in speech movements to arise in this way. It is accordingly quite logical to assume that the trait which appears in this test, in a reaction

which is entirely outside of the field of speech, is an important part of the stutter mechanism. It is probably a general characteristic of the behavior of the individual, and common to many other than vocal reactions. This accords with the known fact that spasms and blockings analogous to stuttering occur in other than oral expressions, e. g., in writing movements.

The conclusions which at present seem most evident from these results may be formulated briefly as follows:

1—No reliable evidence is found in the tapping test for any defect of coordination or motor control at the level of the rapid but coarse muscular movements of tapping.

2—Rapidity in the type of reaction involved in tapping appears to be associated with recovery from stuttering.

3—Regularity of tapping rate seems to be associated with improvement in stuttering.

4—In the hand coordination test stutterers, in most groups, tend to do relatively better when following the beating of a metronome than when they simply go as fast as they can, while the opposite appears in normals. These findings are in harmony with the well known effect of imposed rhythm on the stutterer's speech. Here it apparently plays an important rôle in reactions entirely outside of the vocal field.

5—The stage of improvement in stuttering seems essentially to involve a disturbance in the types of coordination required in the hand coordination test.

6—The stutterers have in the block test a distinctly poorer memory span for movements than normals. It seems logical to consider the defect or difficulty to be one of coordination of the kinaesthetic and other imaginal elements when they appear in large numbers simultaneously, which happens, e. g., in conscious speech movements.

7—No evidence is found in the two-dot test for any weakness in the kinaesthetic field, that is, that part which is involved in arm movements.

8—The results from the visual imagery tests (from 12 subjects) show more relation to intelligence than to severity of stuttering.

9—No reliable difference is found between stutterers and normals as to speed or as to number of false reactions in the complex reaction-time test.

10—Stutterers do not show any more instability or variability of reactions in the complex reaction-time test than do normals. At least in this type of situation, stutterers appear to be fully as stable in their reactions as normals are.

11—Ex-stutterers are unusually regular in speed in the com-

plex reaction-time test. This trait is evidently associated with recovery from stuttering.

12—Stutterers do not show in the inhibition test any tendency to differ uniformly from normals in mere amount of inhibition.

13—Stutterers have in the inhibition test fewer partially inhibited responses than normals. It seems very probable that a general lack of ability to inhibit an impulse after it has found partial expression is an essential factor in stuttering.

## AN EXPERIMENT IN THE USE OF MUSIC IN AN ARCHITECTURAL DRAFTING ROOM

By ESTHER L. GATEWOOD

This study undertakes to investigate the use of music in a specific situation, namely in an architectural drafting room, when the men are at work. Is the use of music of any value or significance in such a situation?

Many old legends relate stories of the effects of music upon the feelings and actions of the listener. More recent literature, with some historical authority, reports similar instances. Recently, experimental studies have been undertaken to scientifically prove or disprove these beliefs. Practically all of the reports and most of the experimental work have dealt with the passive listener, from the standpoint of enjoyment, of musical pleasure.<sup>1</sup> The effect of music upon work produced or upon people at work has not been systematically studied, although occasional references to its possible effect have been made.

A few scattered unregulated observations have been made on the use of music in learning typewriting, and certain other mechanical processes. These have, however, never been done under laboratory conditions, where the situation is controlled. No objective measurements have been made, nor have there been organized introspective reports. The technique and the reliability of the results are open to question. Attempts to obtain definite details from those who are reported to have made such experiments have resulted in the receipt of letters proclaiming the "wonderful effects," "greatly stimulating effect," "the unbelievable aid," etc., but giving no quantitative data.

For several months, my office was located, by chance, near the large drafting room of the department of Architecture of the Carnegie Institute of Technology. The work of the

---

<sup>1</sup> Gehring, Albert. *The Basis of Musical Pleasure*. 1910, New York.  
Gilman, B. J. Report of Experimental Test of Musical Expressiveness. *Am. Jour. Psych.*, 1891, 1892, vols. 4, 5.

Downey, June. A Musical Experiment. *Am. Jour. Psych.*, 1899, v. 9.

Weld, H. P. An Experimental Study of Musical Enjoyment. *Am. Jour. Psych.*, 1912, vol. 23.

Mohler, Louis. *Music Moods*, 1916. Columbia Graphophone Co., N. Y.

students in this department is so arranged that each is given a big problem every six weeks. Most of this time they spend studying out the various parts of their respective problems. All the details are carefully planned and tried out before actual work on the drawing is begun. About a week before the finished sheet is due, they start to work.

Toward the end of the period, the men work early and late. The last four or five days, the students are "on charette."<sup>2</sup> The last few nights of the charette period, they work until very late, two A. M. being the closing time of the longest evening allowed.

Several times, when students were "on charette" some of the men had asked me for the use of the phonograph and a few records. Two of the men usually picked out a group of selections, which they used. The exact content of these groups, I do not know, except that a large proportion of it was late popular music, much of it jazz.

There was no question as to the enthusiasm of the men over having music occasionally. I was told frequently by various ones how much they had enjoyed it and how much it had helped them. However, from such remarks and testimonials, it is both unwise and unsafe to draw conclusions. That is exactly the unreliable basis for much of the popular literature on the influence of music. The fact that a few liked it was no evidence that others did, nor was there any certainty that the selections which these men chose represented the type of music others would like. Nor could one conclude that the selections which were chosen were the best for this particular purpose. Their reports were suggestive, but not conclusive, and I undertook to study these questions:

(1) Is music of any help in this particular situation, namely, the actual work of completing plans and drawings in an architectural drafting room?

(2) Is music, on the contrary, an unfavorable distraction to the worker?

(3) What type of music is most preferred under such circumstances?

(4) If music is desirable, in what way does the worker feel that the music helps?

---

<sup>2</sup> The term, "on charette" is brought over from the French. When plans and charts were due, a little cart used to go about through the streets to gather them in. Draftsmen were often seen running along behind the cart putting on the finishing touches. The term is now used to indicate the period of intense work, when finishing architectural plans or drawings.

(5) If music were constantly available, in what time units would it be used to the best advantage?

The situation in the drafting room made an ideal one for experimentation. Arrangements were made for one of the assistants to have the instrument brought into the drafting room just as on the previous occasions when they had borrowed it. They kept the instrument going in the usual manner, except that the re-creations used were those which I had selected for the occasion. The experimenter was merely a visitor with some of the students and none of them knew until afterwards that the music was being played with a purpose on this occasion. Many of the men expressed surprise when they realized why the writer had been there. The music was therefore introduced in a manner which caused no disruption to the work, and none of the antagonism to experimentation, which is so often felt, was aroused. The conditions were carefully controlled and at the same time, were perfectly normal so far as the students were concerned.

The data sheet, like the accompanying form, was prepared. This was so planned that it was unnecessary for one to write lengthy answers to any of the questions; a check mark or figure sufficed. This was done for two reasons. First, it made the answers comparable. Too often results of such experiments are given in terms of the remarks of the observers, and there is no satisfactory means of converting the opinion of one into the same terms as the opinion of another. Second, it required as little time as possible from the men at work, and their serious coöperation was thus obtained, whereas, with a lengthy questionnaire which required much of their time, they would have been unwilling observers. These report sheets were passed out at the end of two hours. Only then did the men suspect that the music had been given to them for a purpose. They were much interested and serious in their reports.

An Edison "laboratory model" phonograph, the same which had been loaned to them on previous occasions was used. A set of re-creations (the Edison trade name for their phonographic records) was purposefully selected so that it included various types of music, of instrument and rendition. The entire group contained forty-five selections, several each of band, orchestra, chorus, instrumental solos and vocal solos, rhythmic waltzes, one vocal quartet and many jazz or popular dance numbers. Some of the jazz selections were not used, so that the number was somewhat reduced. The instrumental solos included violin, piano, flute and cello; the vocal solos,



each kind of voice, soprano, contralto, tenor and baritone. I had previously arranged the re-creations in such a way that there was a fairly even distribution of each type.

Name.....Date.....  
 Check the appropriate answers in the blank spaces indicated.  
 Does the presence of music make your work any easier?  
     Yes.....No.....  
 Does the music distract you from your work?  
     Yes.....No.....  
 Check the types of music which you liked best (not in general, but with your work tonight). Check one of each pair.  
     Instrumental.....Vocal.....  
     Loud.....Soft.....  
     Slow.....Fast.....  
     Non-jazz.....Jazz.....  
     Familiar.....Unfamiliar.....  
 Number in order the types you preferred most with your work.  
     Standard orchestral selections.....  
     Instrumental (violin, piano, etc.) solos.....  
     Vocal solos.....  
     Jazzy dance numbers.....  
     Rhythmic waltzes.....  
     Lively chorus numbers.....  
 How do you think music helps? Check.  
     Rhythm speeds actual work.....  
     Keeps you happy.....  
     As rest between work.....  
 Give any other reason.....  
 Assuming you could have music whenever you wanted it, how often would you have it? Check.  
     Rarely and for a short period.....  
         "                    " long.....  
     Often and for a short period.....  
         "                    " long.....  
     Most of the time.....  
     Not at all.....

Remarks.....

*Results.*—Replies were received from fifty-six men, although not all answered every question. To the first question, "does music make work easier?" forty-nine (49) answered 'yes' and only five (5) 'no.' Six (6) men declared that music distracted them from their work, while forty-seven (47) answered that it did not. The seeming inconsistency of the one person who claimed that music distracted him and yet that it made work easier is not impossible. He was one of those who felt that music was a help if used as a rest at intervals between work periods.

In answer to the third question, on the various types of music most suitable, the preferences are expressed in the following table, which is arranged to show the contrasting types:

instrumental.....	37	vocal.....	3
loud.....	19	soft.....	19
slow.....	15	fast.....	12
non-jazz.....	15	jazz.....	22
familiar.....	30	unfamiliar.....	2

Not only do the figures show that instrumental selections are very much preferred to vocal numbers, but also the comments made whenever vocal selections were played, showed that they were irritating. According to a questionnaire sent out to a large number of phonograph owners, many more people express a liking for vocal music in general than for instrumental. In this sort of a circumstance, however, when the individuals are working, and the music is merely an incidental feature, instrumental music is greatly preferred. Either voice quality, or what is more likely, the presence of words, proved to be an irritation as well as a distraction. The effect is very like that which one experiences when he is trying to prepare a lesson or write a theme, and some one in the room talks to himself or to some one else. There is usually an effort to listen to at least a part of the words, and the attention of the worker shifts back and forth from the song to his work. This interruption is not evident with instrumental music.

Vocal music compels the attention, whether willing or unwilling, of the audience. It forces itself upon its hearers. This fact is evident in a drawing-room party. If some one renders a vocal solo, the audience usually gives attention whether the performance be good or bad. On the other hand, if some one gives a piano solo, oft-times in spite of the fact that it is unusually well done, the music is interpreted as a signal for conversation. Certainly, if the audience is given to conversation, the presence of instrumental music is not the distraction that vocal music is.

There was an almost equal division of opinion on the question of tempo and volume. Usually slow tempo and soft were preferred by the same individual and fast tempo and loud by the same individual. It is difficult to determine just how much of this is indicative of the individual's own personal preferences, and his preference for this particular kind of music when working. Special pains were taken however to have the comparison on the basis of the selections just heard, and

I do feel that the judgments rendered are reliable. One can readily understand why loud and fast, and soft and slow are linked up together. Loud slow music forces itself upon the listener, in accented punctuated measure that distracts, and also causes the movement to lag. Slow soft music, on the other hand, is very soothing and rarely annoys or even distracts the hearer.

More scored jazz than scored non-jazz, and several did not score either. Those who did not score either gave answers to some of the other questions which would make a scoring on jazz very unlikely, so that the division between the two is approximately equal.

Familiar music is greatly preferred to unfamiliar music. Several men explained that if the selection were a new one, they wanted to listen to the melody, and if it were a song, they listened to the words. In a selection that is already familiar, little or no attention is consciously given to the words, and the music is merely a pleasant accompaniment to the work, and in no way a distraction.

The selections used for the experiment included several of each of the following types: standard orchestral, instrumental solos, vocal solos, jazzy dance numbers, rhythmic waltzes, and some lively chorus numbers. In question four, the men were asked to indicate the order in which they preferred the several types. The relative order of the various types proved to be the following:

1. Jazzy dance
2. Standard orchestral
3. Rhythmic waltz
4. Instrumental solo
5. Lively chorus
6. Vocal solo

This rank order is based on the number of first, second and third rank votes given each type by the entire group of men. From whatever angle the data are studied, *solo numbers* are less desirable than ensemble, and vocal music less desirable than instrumental. It is evident that *vocal music* is not suitable for such a purpose as we are investigating. Those who gave the most careful consideration to the whole questionnaire, ranked instrumental solos and orchestral numbers higher than they did the two dance types. This however may be an indication of the more serious type of individual and his preference for music other than the dance.

One other argument, from an objective viewpoint, lends favor to the use of standard orchestral selections in preference

to the jazz type. The room is noticeably noisier when loud jazzy numbers are played and more talking goes on. Some of the energy of the worker goes out in a form that does not show in his drawings. When quieter, more standard selections are played, there is less confusion in the room and one worker does not disturb another. It is possible that some individuals may work better in short periods between jazz numbers, and so in the end accomplish as much as those who work continuously using a different type of music. This is not likely of the group as a whole, although a certain amount of the livelier sort of music is doubtless useful.

Determination of the causes of the beneficial results from using music during the charetting period is more difficult. A comparison of the introspective judgments of the entire group brought out some interesting facts. Only three explanations were suggested on the data sheet, no others having at any time been offered. Asking on the data sheet for other reasons, and questioning some of the men informally afterwards brought out no further explanations.

Twenty-two (22) of the men felt that the music actually speeded their work, that fast rhythmic selections quickened their movements. Some explained that they felt a tendency to keep the rhythm of the music as they drew, as for example, in shading. However, with most of them, the music seemed to be rather a nervous stimulation, which brought out increased activity in the work. This speeding the movements is in keeping with the preference on the part of many for the quick jazz music, and the livelier orchestral selections.

Forty-one (41) of the draftsmen asserted that music keeps them happier, and that they can do more work and better work when they are in good spirits. Twenty (20) of them explained that the music furnished a rest period between working periods. This function may not be unlike that of keeping the men happy, although the temporal use is different. To those who prefer it as rest, at intervals, there arises the difficulty of determining the proper intervals, and indeed this element is apt to vary with individuals to such an extent that with a large group it would be extremely difficult. However, what seems to be the explanation is that the listener is able to ignore the music and consider only his problem when he is working and really notice the music only when he pauses to rest or to think over the next step in his work.

Not a few of the men indicated more than one of the above replies. Many of them suggested that the music helped keep them in good spirits, and also that it served either as an actual

accelerator to their work, or as a rest. There is certainly no doubt in the minds of the men who have tried it that either by actually speeding the movements of the worker or by mentally stimulating him, music does increase the ease with which this work is done.

As a result of the last question, concerning the most satisfactory time distribution, the desirability of the music was further substantiated. Twenty-one (21) replied that if they could choose the time, they would have the music often and for short periods, nine (9) often and for long periods. Twenty-eight (28) answered that they would like it most of the time. Some few stated explicitly that they would like it most of the time, but in short periods, some others that their preference would vary with the circumstances. Only one (1) asserted that he would want it "not at all," none "rarely and for short periods." No exact definition was given of the length of a long period or a short period. A long period was taken to mean continuous playing for approximately half an hour, while a short period was taken to mean the playing of three or four records, followed by a period without music. Many of the men expressed the opinion that they would prefer it frequently and for short periods, while an equal number prefer it most of the time. There seems to be no doubt in the minds of these men at least that music is a highly desirable accompaniment to their work.

#### CONCLUSIONS

(1) The use of music, rendered by means of the phonograph, was felt by the workers in an architectural drafting room to be of decided advantage.

(2) Only a small percentage of the workers reported that they found the music a distraction or a hindrance to their work.

(3) Instrumental music is much preferred to vocal music for this purpose.

(4) Music unfamiliar to the listener is not as desirable as familiar music. It becomes more or less of a distraction.

(5) According to the reports of the workers, the aid of music is of two sorts. Many feel that it actually speeds movement, and practically all the workers find that it has a beneficial effect on the mood or spirits, which in turn is reflected in the work. It is not unlikely that it performs both of these functions.

(6) Music is not a feature to be used rarely as a sort of diversion or intermission, but instead may be used to advan-

tage along with the work, a large part of the time. To have music frequently and for short periods appears to be most desirable and beneficial.

So many unfounded statements have been made on the advantages or values of music in industrial situations, that the natural result has been a tendency to consider all such reports the mere opinion of some sentimental enthusiast. It is, therefore, of peculiar significance that the results of such an experiment, unbiased by any popular enthusiasm, should be positive, and that the workers themselves should be so definitely convinced of the desirability of the use of music. It is not safe to generalize from this particular performance to other industrial situations, but it is not unlikely that experiments in many other fields may show similar results. However, we can safely say that the introduction of certain kinds of music proves to be both desirable and beneficial in the drafting room.

## THE RESULTS OF SOME TESTS ON FULL AND MIXED BLOOD INDIANS<sup>1</sup>

By THOMAS R. GARTH, University of Texas

The objective of Racial Psychology of an experimental character should be first to find norms and the measures of variabilities around these norms. These should stand as inventories of the races measured. But if we follow out the latter conception the complete distributions of the measures would be a truer representation of the facts. And having obtained these, our next intention would be to make comparisons between the races. Even though our method may be avowedly scientific, we may well take caution in the making of these comparisons because of the ever present bias due to race prejudice which may cause us to hasten to conclusions and thus endanger the so-called inferior races with the stigma of being rated low. For this reason there ought to be a canon in the study of *race* psychology similar to the famous Canon of Lloyd Morgan in his studies of animal behavior. While Morgan's basal principle reads: "In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty" (of lower animals), "if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale;" the canon for *race* psychology might read after this fashion (with due apologies to the British scientist): In no case may we interpret an action as the outcome of the exercise of an inferior psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands higher in the psychological scale, but is hindered by lack of training.

There is still a third objective in the study of racial psychology of which little has been said and of which insufficient use has been consciously made. Mental measurement has taught us to seek out the subnormal child so that we may give him such opportunity as he may need to bring him to his full development. Measures of the inferior races have indicated some superior individuals in the racial group who even take a position beside those above the median of the superior race with which we are comparing them. The third objective then of racial psychology should be to discover these superior individuals who find themselves contending with unfavorable environmental and traditional conditions. It

<sup>1</sup> Paper read before American Psychological Association, Chicago, December 30, 1920.

is not in the province of racial psychology to indicate what should be done for these "superiors," but only to find these individuals.

While the problem involved in the experiment which is to be described and interpreted is to find out how Mixed and Full Blood Indians differ in the results of their performances of nine psychological tests, it has all the objectives mentioned above embodied in it, though it arrives at them in a preliminary fashion.

The subjects included three hundred and eighty-four Indians of the United States Indian School at Chilocco, Oklahoma, of whom 198 were males, 77 being mixed blood and 121 full blood Indians, and 186 females, 78 being mixed and 108 full bloods. Nearly all of them were Plains and Forest Indians. Their ages ranged from 9 to 26 years and their educational attainment ranged from the fourth to the tenth grade. To these Indian students were given four association tests: Opposites, genus-species, part-whole, and free (continuous) association; three memory tests: concrete and abstract rote memory and logical memory; and two word-building tests: *apirle* and *acobut*.

These tests were given in the spring of 1920. The technique was the same as that used by Pyle and his associates in giving all of these tests to Whites and Negroes, and some of them to Chinese. (School and Society, Vol. I, p. 357 and Vol. II, p. 264.) Other tests were given but the results are not presented in this paper.

While it will be observed from the above description that we have here groups of different ages, education, and sex as well as different blood, the subjects were classified with reference only to sex, racial blood, and age. A further breaking up of the original group of subjects into educational grade groups would make the sub-groups too small, so that this classification has been deferred until larger numbers of Indian children can be secured for testing with the same material.

The first draft of the data is presented in the accompanying tables, the data for the males being given in Table I and that for the females in Table II. In both tables, as we have said, the data for Mixed Bloods and Full Bloods are kept separate and these large groups are again subdivided into age groups running from 9 years to adult age, which means here from 19 to 26 years. The number of cases for each age group is designated also. In the Mixed Blood groups for both sexes the average degree of Indian Blood for an age group is indicated, and for all age groups we have the average school grade attainment. The subjects were ju



finishing the designated school grades, for the tests were given in March or toward the close of the school year. The tables give the averages, the A. D., P. E., and range for all age group divisions for every measurement undertaken. In addition is shown the overlapping in each age group from the 14th year up, which indicates the per cent of Mixed Bloods attaining and exceeding the median score of the Full Bloods.

The norms for Mixed and Full Blood Indians shown here are of necessity only tentative because of the size of the groups. It is likely that radical changes will occur in them when more racial material of a similar kind can be obtained. However, within the limits of the data at hand, we may say what follows. In our search for differences which these data show between Mixed and Full Bloods of the Indian race, we shall consider these facts:

I. The average of each age group for Mixed Bloods and Full Bloods;

II. The instances where 60% or more of the Mixed Bloods attained the median of the Full Blood scores in each age group; and

III. A comparison of range of the Mixed Bloods and of the Full Bloods.

IV. However, inferences drawn from these findings must be modified by the differences in educational attainment and social status between the Mixed and Full Bloods.

If we take for purposes of comparison the average performance of each group of mixed and full blood individuals for each test we find that the Mixed Bloods excel the Full Bloods in most of the cases. In the case of the comparison of the results for the males, the Full Blood fifteen year and adult groups excel in free association; their fourteen and eighteen year groups excel in concrete memory; their seventeen year group in abstract memory; their fifteen, eighteen year and adult groups in both word-building tests; and their seventeen year group in the second word-building test. That is, there are twelve instances in which the average of these full blood Indians excel out of fifty-four possible instances, or 22% of the time. The case of the females is not even so good as this, for there are only four instances in which the Full Bloods excel the Mixed Bloods, *i. e.*, the adult group in genus-species, the fourteen year and adult group in concrete memory, and the adult group in abstract memory. This makes four out of a possible fifty-four, or 7% of the instances. Then the evidence from comparison of the averages favors the Mixed Bloods.

A fairer test would be to find the overlapping, as for instance, to what extent the Mixed Bloods excel or attain the median of the Full Bloods. We find that for each test 60% and better of the Mixed Bloods attain and exceed the median of the Full Bloods for each age group in all tests except in the following instances: males, fifteen year olds in opposites, eighteen year olds and adults in genus-species, eighteen year olds in concrete memory, seventeen year olds in abstract memory, fifteen year olds and adults in logical memory, fifteen year olds and adults in part-whole, likewise, and fifteen, eighteen year olds and adults in both word-building tests; females, adult group in genus-species, fourteen year olds in free association, and adults in abstract memory, seventeen year olds and adults in part-whole association and also in the first word-building test, and sixteen year olds in the second word-building test. To sum this up the Mixed Bloods have this 60% and better of their respective groups attaining and exceeding the median of the Full Blood in 39 out of a possible 54, or 72% of the time for the males, and 46 out of 54 times, or 84.4% of the time for the females. This likewise favors the Mixed Bloods.

Attention should be called to the fact that in making the above comparisons certain age groups of the Full Bloods stand out prominently in several of the tests, particularly the fifteen year, eighteen year, and adult groups of males, and the adult females. There are evidently more superior individuals in these age groups than in the other age groups of the Full Bloods. This naturally brings us to a consideration of the range in the various age groups for purposes of comparison. Upon making such an investigation we find for the males that the upper range of the Mixed Bloods is greater than that of the Full Bloods 51.8% of the time, using the age groups here as bases for comparison as before, or 28 out of 54 times. For the rest of the time the latter excelled the former in going beyond their upper range, except twice, when their upper ranges were equal. In other words, for every excellent performance of any Mixed Blood individual, nearly half the time there was a full-blood Indian to match his performance or to excel it. It is not so for the females of this blood group, for when their upper range is compared with that of the other blood group of their sex it is seen that 40 times out of 54 the Mixed Blood upper range is greater than that of the Full Bloods i. e., 74% of the time. In only nine instances was the upper range of the latter female group greater than that of the former and in five instances it was equal to it.

As to the Full Blood lower range, it was below that of the Mixed Bloods 34 out of 54 times —62.9% of the time for the males, and it was about the same, 35 times out of 54 —64.8% for the females, and as good or better for both sexes during the rest of the time. The presence of the superior performers among Full Bloods in the 14, 18 year, and adult groups for the males is again made evident by this examination of the range, and it is somewhat so for the female adults.

Before undertaking to draw conclusions from the above comparisons, two additional factors should be considered, *i. e.*, social status and education. The data at hand offer little information with reference to social status. We can say that the males, from the fourteenth year up, were on the average 54% Indian blood, and the females slightly more, but both are practically the same, and while we know little of the social status it is safe to draw the inference that since they were nearly half white, all of them being mixtures of whites and Indians, the influence of the white parent would necessarily improve the social status from the white standpoint and this would be above that of the individual of full blood parentage more or less biased by primitive traditions. And as to education, the average attainment for the Mixed Blood males is 7.4 grades, ranging from 5th to 10th grade and for the females was 7.3 grades, ranging from 4th to 10th; whereas for the Full Blood males the average school attainment is 6.2 grades ranging from 4th to 10th grades, and for the females 5.8 grades, ranging from 4th to 9th.

The relative lack of social pressure to go to school is evident in the Full Blood group which, to be sure, is the result of primitive apathy or lack of enthusiasm for the white man's education. Nevertheless, this should not be regarded as peculiarly an Indian trait, though it assuredly militates here against better performance in the psychological tests for the Full Blood Indian.

From the above experiment then we conclude that in a comparison of Mixed and Full Blood Indians as to their ability in the performance of certain tests, the Mixed Bloods tend to excel the Full Bloods on a score of averages, measure of per cent of the former to attain the median of the latter and as to upper range of scores, but many Full Blood individuals attain and excel the performance of Mixed Blood individuals. Furthermore the scores of the Mixed Bloods is favored by their superior social status and educational opportunity. However, the writer doubts if an equality of school attainment would remove the indicated differences, not as measured, but in their trend.

## FROM THE PSYCHOLOGICAL LABORATORY OF THE UNIVERSITY OF TEXAS

## THE RESULTS OF SOME TESTS ON FULL AND MIXED BLOOD INDIANS

The problem is to find out how Mixed and Full Blood Indians differ in the results of their performance on nine psychological tests.

TABLE I

Males:		9	10	11	12	13	14	15	16	17	18	Adult
Age—Years.....												
Mixed Bloods:												
No. Cases.....		1	0	3	2	5	7	11	13	9	12	14
Av. De. I. B.....		75		.50	.38	.75	.52	.61	.57	.47	.59	.47
Av. Years Ed.....		4	0	.47	4.0	5.2	5.1	6.0	7.7	6.6	7.9	7.8
A. D.....							1.5	.7	1.1	.8	1.3	.9
P. E.....							1.2	.5	.9	.6	1.0	.7
Range.....		4		4-5	4-4	5-6	5-6	5-8	5-10	5-9	5-9	5-10
Full Bloods:												
No. Cases.....					4	5	10	10	24	17	22	29
Ave. Years Ed.....					5.0	4.6	4.8	5.4	5.6	5.9	6.5	7.6
A. D.....							1.0	1.1	1.7	1.3	.9	1.9
P. E.....							.8	.9	1.4	1.0	.7	1.4
Range.....					4-7	4-5	4-6	4-8	4-8	4-9	5-9	4-10
Opposites' Test:												
†M. B. Ave. Score.....		10		7.3	7.5	11.2	9.6	9.2	12.0	12.9	11.0	11.4
A. D.....				.8	.5	2.3	2.7	3.6	4.6	2.8	3.8	3.4
P. E.....							2.1	2.9	3.6	2.2	3.0	2.7
Range.....				6-8	7-8	9-17	4-15	0-16	1-20	7-19	2-18	5-15
F. B. Av. Score.....					8.0	8.8	8.7	9.0	9.3	6.9	8.7	9.3
A. D.....					3.5	1.5	1.8	2.6	3.2	3.0	2.3	3.7
P. E.....							1.4	2.1	2.5	2.3	1.8	2.9
Range.....					3-15	5-11	2-17	4-14	3-15	3-12	3-17	1-19
††%M. B. at Med. of F. B.							62.5	55.5	69.2	100	72.8	75.9

\*Average degree of Indian Blood.

F. B. indicates Full Bloods.

†M. B. indicates Mixed Bloods.

††Per Cent. of Mixed Bloods attaining medium of Full Bloods.

Genus-Species' Test											
M. B. Av. Score.....	12	6	.6	7.6	8.4	9.9	11.7	10.4	10.1	10.0	
A. D.....	..	.5	2.2	2.1	1.8	3.7	3.2	3.5	3.3	3.3	
P. E.....	..	..	..	..	1.4	2.9	2.5	2.8	2.8	2.7	
Range.....	..	4-8	2-10	3-10	5-10	3-16	4-20	4-19	4-17	5-19	
F. B. Av. Score.....	..	..	6.8	6.2	7.4	7.4	7.7	9.9	9.0	9.5	
A. D.....	..	..	2.2	2.2	4.0	2.2	3.2	2.7	2.9	4.0	
P. E.....	..	..	..	3.1	1.8	2.5	2.1	2.3	2.3	3.2	
Range.....	..	..	5-11	3-9	1-19	4-13	1-15	4-13	2-16	3-16	
Per Cent. of M. B.'s	..	..	..	..	..	..	..	..	..	..	
Att. Med. of F. B.'s.....	..	..	..	..	85.7	63.6	83.7	62.5	50	57.1	
Part-Whole Test:											
M. B. Av. Score.....	12	7.5	6.0	7.8	9.2	10.8	11.8	10.9	11.9	10.2	
A. D.....	..	?	?	?	2.2	4.5	3.0	2.3	5.2	5.2	
P. E.....	..	..	..	..	1.7	3.8	2.5	1.9	4.1	4.3	
Range.....	..	5-9	..	5-11	4-13	0-19	5-18	5-17	3-19	1-20	
F. B. Av. Score.....	..	..	7.7	7.0	8.2	9.6	8.9	9.0	9.0	9.4	
A. D.....	..	..	?	?	2.2	2.1	4.5	3.6	3.6	4.4	
P. E.....	..	..	..	..	1.9	1.7	3.7	2.9	2.7	3.4	
Range.....	..	..	1-19	3-9	4-12	0-13	0-20	0-14	0-20	1-19	
Per Cent. M. B.'s Att.	..	..	..	..	..	..	..	..	..	..	
Med. of F. B.'s.....	..	..	..	..	80	55.6	75.0	60	60	57	
Free Association Test:											
M. B. Av. Score.....	37	49.3	56	50	46.7	44.8	50.3	46	50.4	44.3	
A. D.....	..	..	7	6.4	7.7	9.9	10.9	5.2	8.8	9.6	
P. E.....	..	..	..	..	6.2	8.1	8.7	4.1	7.1	7.7	
Range.....	..	42-55	49-63	42-60	32-58	8-72	29-66	23-59	33-70	16-78	

FROM THE PSYCHOLOGICAL LABORATORY OF THE UNIVERSITY OF TEXAS—Con.  
THE RESULTS OF SOME TESTS ON FULL AND MIXED BLOOD INDIANS

Males:

Age, Years.....	9	10	11	12	13	14	15	16	17	18	Adult
F. B. Av. Score.....	..	..	..	49.8	51.6	42.0	45.9	42.5	39.3	44.4	44.3
A. D.....	..	..	..	8.8	5.3	9.4	12.5	10.5	10.9	8.0	11.5
P. E.....	..	..	..	..	..	7.6	9.8	8.5	8.7	6.1	8.9
Range.....	..	..	..	35-67	42-58	28-58	27-72	11-67	14-70	5-75	10-84
Per Cent. of M. B.'s.....	..	..	..	....	....	..	..	..	..	..	..
Att. Med. of F. B.'s.....	..	..	..	....	....	71	54.4	62	78	75	54.5

Rote Memory Test—Concrete:

M. B. Av. Score.....	44	38.2	35.5	39.8	41.4	42.6	41.2	40.5	40.5	43.9
A. D.....	..	..	..	..	3.5	5.8	5.6	5.6	3.5	4.2
P. E.....	..	..	..	..	2.8	4.7	4.2	4.5	2.7	3.5
Range.....	..	33-45	33-38	33-44	36-51	27-55	30-52	31-46	34-47	33-59
F. B. Av. Score.....	..	..	35	33.8	37.4	41.8	39.9	38.8	40.8	39.3
A. D.....	..	..	..	..	2.8	2.7	3.6	7.2	4.5	1.4
P. E.....	..	..	..	..	2.2	2.4	2.8	5.8	3.5	1.2
Range.....	..	..	23-42	24-39	33-42	36-47	31-49	26-49	28-56	32-53
Per Cent. of M. B.'s.....	..	..	....	....	..	..	..	..	..	..
Att. Med. of F. B.'s.....	..	..	....	....	62	70	67	88	54	92

Rote Memory Test—Abstract

M. B. Av. Score.....	36	34.5	26.0	33.0	32.8	33.4	37.7	30.9	38.8	37.2
A. D.....	..	..	..	..	4.9	3.2	7.6	5.1	3.9	5.4
P. E.....	..	..	..	..	3.8	2.7	6.3	3.8	3.1	4.3
Range.....	..	31-37	24-28	24-49	15-39	21-43	19-49	22-42	24-46	26-48
F. B. Av. Score.....	..	..	29.5	27.6	26.1	33.1	30.4	31.1	33.8	33.6
A. D.....	..	..	..	..	7.1	5.5	4.9	7.9	7.4	5.2
P. E.....	..	..	..	..	6.1	4.7	3.8	6.3	6.2	4.3
Range.....	..	..	22-37	25-29	17-43	19-45	18-44	6-43	17-52	21-44
Per Cent. of M. B.'s.....	..	..	....	....	..	..	..	..	..	..
Att. Med. of F. B.'s.....	..	..	....	....	87	70	75	50	85	69

Logical Memory Test:										
M. B. Av. Score.	36	21.3	35.0	25.0	28.9	21.5	30.3	34.9	24.5	26.2
A. D.	...	5.8	7.0	10.4	7.2	11.6	8.6	10.1	4.6	8.7
P. E.	...	...	...	...	5.8	9.4	6.8	9.9	3.6	6.9
Range.	...	9-25	28-42	15-43	9-39	0-47	11-43	0-50	9-31	12-43
M. B. Av. Score.										
A. D.	...	...	24.6	14.5	20.4	19.9	21.9	21.2	22.0	22.6
P. E.	...	...	7.9	8.0	13.1	14.3	12.2	10.1	10.7	10.0
Range.	...	...	15-35	0-22	0-35	8-34	4-37	0-44	8.6	8.2
Per Cent. of M. B.'s	...	...	...	...	...	...	...	...	0-51	0-45
Att. Med. of F. B.'s.	...	...	...	...	86.0	58.0	82.0	75.0	80.0	53.0
Word Building AEIRLP Test:										
M. B. Av. Score.	9	17	11.0	14.8	14.1	13.0	13.4	13.9	11.5	11.3
A. D.	...	...	...	...	4.1	4.0	4.4	3.9	4.5	4.5
P. E.	...	...	...	...	3.3	3.4	3.5	3.1	3.5	3.4
Range.	...	16-20	10-12	9-17	6-20	4-20	3-20	9-21	0-18	0-21
F. B. Av. Score.										
A. D.	...	...	14.8	11.4	12.4	13.5	11.9	11.3	12.1	12.5
P. E.	...	...	...	...	4.9	3.0	3.3	3.1	3.2	4.9
Range.	...	...	...	...	3.8	2.4	2.6	2.4	2.5	3.9
Per Cent. of M. B.'s	...	...	11-18	5-18	3-22	9-18	2-18	5-18	4-21	0-22
Att. Med. of F. B.'s.	...	...	...	...	75	37	66.7	66.7	45.5	55.5
Word Building AEOBMT Test:										
M. B. Av. Score.	15	13.3	11.0	15.0	13.2	13.0	14.4	12.0	11.6	12.2
A. D.	...	...	...	...	2.7	3.2	3.8	2.4	5.1	4.1
P. E.	...	...	...	...	2.2	2.5	3.1	2.1	1.1	2.0
Range.	...	10-16	11-11	10-21	8-18	5-19	8-19	8-17	1-18	2-20
F. B. Av. Score.										
A. D.	...	...	12.5	9.0	11.3	13.6	12.5	12.3	13.8	13.8
P. E.	...	...	...	...	3.4	3.6	3.8	3.4	2.0	3.7
Range.	...	...	...	...	2.8	2.7	3.2	2.6	1.7	2.9
Per Cent. of M. B.'s	...	...	5-20	5-8	5-17	7-19	3-20	5-19	7-18	5-23
Att. Med. of F. B.'s.	...	...	...	...	50	45.5	72.7	60	50	62

FROM THE PSYCHOLOGICAL LABORATORY OF THE UNIVERSITY OF TEXAS—Con.  
THE RESULTS OF SOME TESTS ON FULL AND MIXED BLOOD INDIANS

TABLE II

TABLE II												
Females:												
Age Years.....	9	10	11	12	13	14	15	16	17	18	Adult	
Mixed Bloods:												
No. Cases.....	0	0	2	1	3	15	9	17	15	11	7	
Ave. De. I. B.....	..	..	.50	.50	.62	.56	.50	.57	.57	.57	.53	
Ave. Yrs. Ed.....	..	..	4.0	6.0	6.0	5.9	7.0	7.1	7.9	8.3	8.9	
A. D.....	..	..	..	..	..	1.1	1.8	1.2	1.5	1.7	1.4	
P. E.....	..	..	..	..	..	.9	1.4	.9	1.2	1.4	1.1	
Range.....	..	..	4-4	6	5-7	4-8	4-10	4-9	5-10	6-10	4-10	
Full Bloods:												
No. Cases.....	..	1	1	3	5	18	15	15	27	16	7	
Av. Years Ed.....	..	4.0	5.0	4.0	5.6	5.2	5.3	5.4	6.4	5.2	7.5	
A. D.....	..	..	..	..	..	1.0	.8	1.5	1.0	1.3	1.3	
P. E.....	..	..	..	..	..	.8	.6	1.2	.8	1.0	1.3	
Range.....	..	..	5	4-4	5-6	4-7	4-7	4-9	4-9	4-9	6-9	
Opposites' Test:												
M. B. Ave. Score.....	..	..	15	20	13	11.1	14.2	13.4	14.4	15.2	13.5	
A. D.....	..	..	..	..	..	2.2	3.5	3.7	4.6	2.4	3.2	
P. E.....	..	..	..	..	..	1.7	2.8	2.7	3.8	1.9	2.7	
Range.....	..	13-17	20	11-16	11-16	2-19	0-19	2-20	2-20	11-19	8-19	
F. B. Av. Score.....												
A. D.....	..	6	10	10.4	7.9	9.7	9.7	8.7	10.7	7.4	10.9	
P. E.....	..	..	..	..	..	3.0	3.2	3.8	3.6	3.9	6.2	
Range.....	..	6	9-11	5-14	1-14	4-17	2.6	3.0	2.9	3.1	5.1	
Mixed Blood Att. Med. of F. B.....	..	..	..	..	..	68.8	100	88.2	71.4	100	75	



Genus Species' Test:										
M. B. Av. Score.....	..	16.5	16.0	12.8	11.0	12.4	13.2	13.2	14.0	12.4
A. D.....	..	..	..	3.4	4.4	2.7	4.7	4.1	4.2	3.2
P. E.....	..	..	..	..	3.5	2.2	3.6	3.4	3.6	2.7
Range.....	..	9-15	.. 12	9-16	4-20	6-17	2-20	3-19	8-20	4-19
F. B. Av. Score.....	..	6.0	8.5	10.6	8.0	9.8	7.8	10.8	8.3	13.6
A. D.....	..	..	..	2.3	2.4	3.3	3.5	4.0	3.1	5.6
P. E.....	..	..	..	..	1.9	2.6	2.8	3.4	2.4	4.7
Range.....	..	.. 6	4-13	8-14	0-15	2-16	1-15	3-20	2-20	11-20
Per Cent. of M. B.'s	..	..	..	..	..	..	..	..	..	..
Att. Med. of F. B.'s...	..	..	..	..	73.8	80	88.2	66.7	100	29
Part Whole Test:										
M. B. Av. Score.....	..	10.5	12.0	9.0	7.8	9.2	9.4	9.5	10.7	9.8
A. D.....	..	..	..	..	4.1	4.4	2.1	3.1	3.4	3.6
P. E.....	..	..	..	..	3.4	3.5	1.6	2.4	2.7	3.0
Range.....	..	8-13	9-15	7-12	1-16	3-17	3-16	0-16	4-17	5-14
F. B. Av. Score.....	..	6.0	7.3	6.4	6.3	6.5	6.1	7.9	4.2	7.0
A. D.....	..	..	..	..	2.5	2.8	3.7	3.2	3.4	3.8
P. E.....	..	..	..	..	2.0	2.1	3.1	2.6	2.8	3.1
Range.....	..	.. 6	3-10	1-10	1-11	3-15	0-12	1-14	0-12	2-12
Per Cent. of M. B.'s	..	..	..	..	..	..	..	..	..	..
Att. Med. of F. B.'s...	..	..	..	..	62.5	67	88.2	54	90	50
Free Association Test:										
M. B. Av. Score.....	..	38	55	43	45.8	48.9	53	52.7	55.5	56.6
A. D.....	..	..	..	7.3	7.8	11.9	10.1	10.6	9.9	12.8
P. E.....	..	..	..	..	6.1	8.5	8.0	8.3	7.4	10.2
Range.....	..	36-36	32-53	26-67	16-67	16-67	16-66	17-70	43-68	31-71

FROM THE PSYCHOLOGICAL LABORATORY OF THE UNIVERSITY OF TEXAS—Con.  
THE RESULTS OF SOME TESTS ON FULL AND MIXED BLOOD INDIANS

*Females:*

Age, Years.....	9	10	11	12	13	14	15	16	17	18	Adult
F. B. Av. Score.....	..	..	48	54	36.8	43	46.9	43.2	42.5	39.2	54.5
A. D.....	..	..	..	..	7.0	9.2	9.6	9.9	4.9	10.4	7.3
P. E.....	..	..	..	..	..	7.1	7.5	7.8	3.8	8.5	5.8
Range.....	..	..	..	..	..	..	..	..	..	..	..
Per Cent. of M. B.'s.....	..	0	48	45-62	22-46	23-58	11-67	16-68	25-65	17-62	44-71
Att. Med. of F. B.'s.....	..	..	..	..	..	46.7	67	93.8	73.3	91	71.4

Rote Memory Test—Concrete:

M. B. Av. Score.....	..	..	39	44	45	37.5	44.8	42.3	41.4	45.1	46.3
A. D.....	..	..	..	0	7.0	6.4	3.4	4.4	5.3	4.3	3.6
P. E.....	..	..	..	..	..	5.2	2.8	3.9	4.2	3.5	2.7
Range.....	..	..	29-49	44	31-54	20-50	37-48	27-49	28-52	38-52	40-50
F. B. Av. Score.....	..	42	38	39	40.8	40.4	39.7	37.7	38.7	36.9	46.5
A. D.....	..	..	..	2	4.2	5.7	3.9	6.4	5.1	5.8	5.2
P. E.....	..	..	..	..	..	4.6	3.1	5.2	4.1	4.6	4.2
Range.....	..	..	38	37-41	37-50	28-58	33-47	31-58	22-48	26-51	40-51
Per Cent. of M. B.'s.....	..	..	..	..	..	..	..	..	..	..	..
Att. Med. of F. B.'s.....	..	..	..	..	78	60	88	93	73	82	71

Rote Memory Test—Abstract:

M. B. Av. Score.....	..	..	33	29	39.5	34.8	41	38.6	38.3	40.3	40.5
A. D.....	..	..	..	..	2.8	7.0	4.4	3.1	5.2	6.1	6.3
P. E.....	..	..	..	..	..	5.6	3.5	2.5	4.2	4.9	4.9
Range.....	..	..	28-38	29	32-46	19-49	32-46	31-47	26-47	30-52	25-54
F. B. Av. Score.....	..	32	18	31.5	35.4	28.9	33.5	30.9	32.3	28.9	41.9
A. D.....	..	..	..	..	7.3	6.3	4.7	6.9	7.5	7.5	5.2
P. E.....	..	..	..	..	..	5.1	3.8	5.9	6.0	6.2	4.3
Range.....	..	..	18	23-40	26-46	15-43	25-51	11-45	10-47	9-47	37-50
Per Cent. of M. B.'s.....	..	..	..	..	..	..	..	..	..	..	..
Att. Med. of F. B.'s.....	..	..	..	..	..	66	100	93	67	100	29

Logical Memory Test:									
M. B. Av. Score.....	24	55	35.5	30.5	34.4	29.6	30.9	30.3	33.3
A. D.....	...	...	...	11.6	4.7	7.2	5.4	8.0	13.2
P. E.....	...	...	...	9.2	3.8	5.8	4.6	6.3	10.6
Range.....	21-50	55	26-44	19-47	18-42	12-46	24-44	21-41	13-49
F. B. Av. Score.....									
A. D.....	11	31.5	26.3	24.0	24.2	21.9	23.9	16.9	27
P. E.....	...	...	...	9.5	5.4	9.0	8.7	9.1	10.3
Range.....	...	...	...	7.6	4.2	7.2	6.8	7.3	7.7
Per Cent. of M. B.'s	11	23-40	15-36	6-46	9-34	2-39	3-41	3-34	9-40
Att. Med. of F. B.'s.....	...	...	...	72	90	81	79	93	67
Word-Building AEIRLP Test:									
M. B. Av. Score.....	7	11	12.5	11.7	15.8	15.2	13.7	16.4	13.9
A. D.....	...	...	...	3.9	3.4	4.6	4.3	3.3	4.5
P. E.....	...	...	...	3.2	2.8	3.7	3.5	2.7	3.8
Range.....	3-11	11	1-20	3-20	9-22	8-22	1-21	8-20	6-19
F. B. Av. Score.....									
A. D.....	12	14.5	13.2	10.0	12.5	12.4	8.8	10.1	13.4
P. E.....	...	...	...	3.5	4.6	5.0	2.5	5.3	2.6
Range.....	...	...	...	2.8	3.7	4.1	2.2	4.8	2.1
Per Cent. of M. B.'s	12	10-19	10-16	1-20	2-20	1-20	0-22	0-18	7-16
Att. Med. of F. B.'s.....	...	...	...	60	66.7	75	45.1	83.3	42.9
Word-Building AEOSMT Test:									
M. B. Av. Score.....	11.0	15.0	12.5	11.6	14.4	14.2	13.8	17.0	13.0
A. D.....	...	...	...	3.9	4.0	3.2	5.0	3.0	3.2
P. E.....	...	...	...	3.2	3.3	2.7	4.3	2.5	2.4
Range.....	11-11	15	3-19	2-18	4-20	4-20	1-21	3-22	2-16
F. B. Av. Score.....									
A. D.....	1.0	15.5	12.8	11.4	11.8	13.1	12.4	12.7	12.7
P. E.....	...	...	...	3.4	3.7	2.6	3.2	4.4	4.5
Range.....	...	...	...	2.8	3.2	2.1	2.6	3.5	3.4
Per Cent. of M. B.'s	11	15-16	10-16	4-17	0-18	2-14	1-19	5-21	2-20
Att. Med. of F. B.'s.....	...	...	...	46.7	70	56.2	66.7	90	85.7

TABLE III

In this table the data of Tables I and II are combined into totals for both sexes and all ages, but the division of Mixed and Full Blood remains.

No. Cases	Age		Educational Attainment		Opposites Test-Score			
	M.B.*	F.B.*	M.B.	F.B.	M.B.	F.B.		
Average.....	158	226	158	226	157	214		
A. D.....	16.4 yr.	16.1 yr.	6.9 yr.	5.9 yr.	12.2	8.4		
P. E.....	1.1	2.4	1.6	1.1	3.4	2.7		
Median.....	.9	2.0	1.3	.9	2.9	2.3		
Range.....	15.7	16.1	6.4	5.3	12.2	8.5		
% Reaching Med. of F. B.....	9-23	10-26	4-10	4-10	0-20	0-19		
	43%		67%		78%			
	Genus Species Test-Score		Part-Whole Test-Score		Free Ass'n. Test-Score		Logical Memory Test-Score	
No. Cases.....	M.B.	F.B.	M.B.	F.B.	M.B.	F.B.	M.B.	F.B.
Average.....	166	223	162	209	151	227	164	214
A. D.....	10.7	8.9	10.0	7.3	48.5	43.9	29.3	21.9
P. E.....	3.5	3.4	3.8	3.8	5.0	10.7	9.2	9.6
Median.....	2.9	2.9	3.2	3.2	4.2	8.5	7.8	8.2
Range.....	10.5	8.1	9.5	6.9	49.6	44.5	29.9	22.1
% Reaching Med. of F.B..	2-20	0-20	0-20	0-20	8-78	1-74	0-50	0-51
	67%		78%		64%		60%	
	Rote Memory Concrete-Test Score		Rote Memory Abstract-Test Score		Word-Building Test-acirp Score		Word-Building Test-acohmt Score	
No. Cases.....	M.B.	F.B.	M.B.	F.B.	M.B.	F.B.	M.B.	F.B.
Average.....	155	224	155	224	153	223	157	224
A. D.....	39.9	39.4	36.6	31.8	13.5	11.6	13.1	12.6
P. E.....	2.7	5.7	5.4	2.9	4.2	3.7	3.9	.9
Median.....	2.3	4.8	4.6	2.5	3.6	3.1	3.3	.8
Range.....	42.1	38.5	36.5	31.5	13.2	12.3	13.4	13.6
% Reaching Med. of F.B..	20-59	22-58	15-54	6-52	0-22	0-22	1-22	0-23
	79%		75%		63%		54%	
*M.B. Mixed Bloods, F.B., Full Bloods.								

\*M.B. Mixed Bloods, F.B., Full Bloods.

## A TEST FOR USE IN THE SELECTION OF STENOGRAPHERS

By M. A. BILLS, Carnegie Institute of Technology

The results of using a battery of tests for the selection of stenographers were described in the September issue of the *JOURNAL OF APPLIED PSYCHOLOGY*. Further data regarding these tests have recently been collected, particularly with respect to the elimination from training courses, of those persons who are certain to fail.

In the previous investigation, the criterion of success consisted of the recommendations of the teachers in English, Typewriting and Stenography, made in January after five months' observation of girls being trained as stenographers. The tests which were found useful in eliminating failures were Bureau Test VI (an 'intelligence' test) and Bureau Test IX (a test of will traits). On the basis of the January ratings as a criterion, a score of 50 in Test VI and a score of 60 in Test IX, Part 10, were assumed as "critical" scores for success in the course. This standard in Test VI would have eliminated eleven failures and one success.

The passing of five more months led to some revision of recommendations by the instructors, and the test scores are, in the present study, regarded in the light of the changes in the criterion. The accompanying figure (Chart I) shows the relation of test scores to the January and the June ratings. Along the ordinate are represented scores in Test VI, and along the abscissa are shown the judgments of success or failure. As the previous article stated, a person who was recommended for continued work in typewriting or stenography, or for advanced courses, was considered a 'success.' Those who were recommended for service in manufacturing or clerical work were considered 'failures.' Each circle on the chart represents a person.

It is apparent that the later and more reliable criterion has justified a readjustment of the critical score for Test VI. The individual scoring 40, who was considered a success in January, has failed to "make good." Of the eight persons scoring between 50 and 60, two were regarded as failures in January and also in June, four were shifted from success in January to failure in June, one considered successful in January had

left before the end of the training course, and only one regarded as successful at the earlier time was still classified as successful.

These data justify the raising of the critical score in Test VI, for success in stenography, ten points—that is, from 50 to 60. In this way, eighteen failures and only one success are eliminated by Test VI alone.<sup>1</sup> In tabular form, omitting the cases of individuals leaving the course before completion, these results are as follows:

Failed in test and failed in course.....	18
Succeeded in test and failed in course.....	9
Failed in test and succeeded in course.....	1
Succeeded in test and succeeded in course.....	39
	<hr/>
Total	67
Per cent correct .....	85

In order to test these results further, three different groups of professional stenographers have recently been given Test VI. The first group consisted of twenty stenographers employed by a single firm. Before the test was given to this group, the stenographers were rated on their value to the firm, by an executive who was in close contact with all of them. Of these twenty stenographers, only one made a score below the critical score set for the students of stenography, i. e., 60, and she was rated lowest by the firm. Table I shows the test scores of the group and their ratings by the executive. The rank correlation is  $+0.75$ . When it is considered that length of service must have played some part in the rating, this correlation must be regarded as remarkably high.

A second group, consisting of seventeen stenographers and eight secretaries employed by a different firm (in an entirely different line of business) were given the same test. They were rated by an executive who knew them all, but did not wish to rank them. He divided them into three groups: those who were "failing," those who were "getting by" and those who were "good." Three of the seventeen were classed as "failures," 4 as "getting by" and 8 as "good." Two had been with the company too short a time to be rated. All of the secretaries were classed as "good." Table 2 gives the ratings, scores and positions of these stenographers.

The mean score for secretaries is 144, for stenographers rated "good," 110; for stenographers rated as "getting by," 65; and for those "failing," 63. Only two made a score less

<sup>1</sup> The standard of 60 in Test IX, Part 10, would have eliminated three who failed in January and none who succeeded. These three persons (indicated on the chart by the circles with dots) left the training course before the June rating.

# TEST FOR USE IN THE SELECTION OF STENOGRAPHERS 375

than the critical score determined by the training school experiment, and neither of these was rated "good."

The third group examined consisted of fourteen high grade stenographers from five different companies. Each company was asked to have the test given to their three best stenographers. The lowest score made was 109 and the mean score was 153.

The data for the original training school group, and for each of the three supplementary groups tested, are given in Chart II. Scores in Test VI are represented by the abscissa, and the number of cases is shown by the ordinate.

From the investigation as a whole, evidence is found that there is a positive relation between mental alertness as indicated by Test VI, and ability in stenographic work. The designation of a score of 60 in this test as a "critical score" for selecting stenographers is justified by the data secured from these various groups.

CHART I.

RELATION OF SCORE IN TEST VI TO SUCCESS AND FAILURE IN STENOGRAPHY

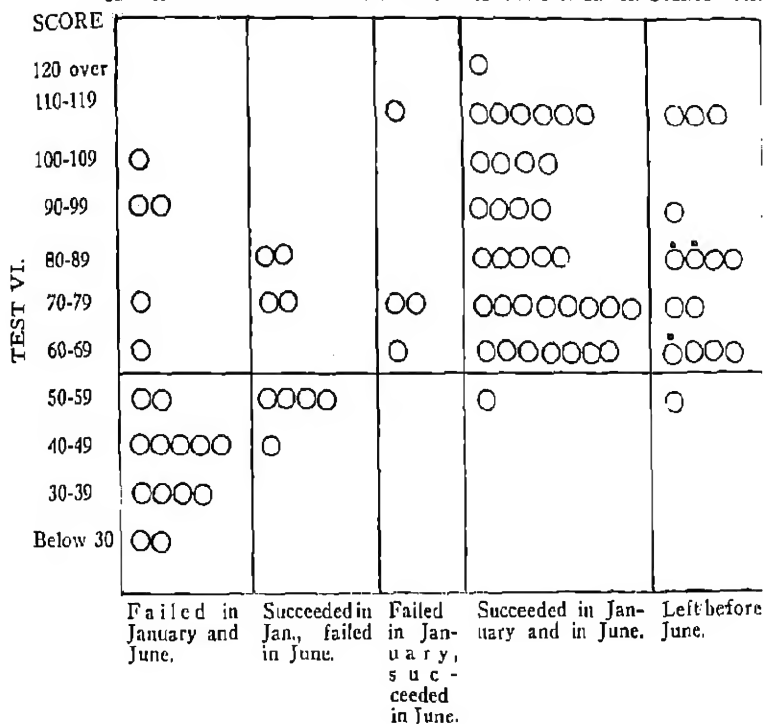


TABLE I  
RELATION OF SCORE IN TEST VI TO VALUE TO FIRM OF TWENTY  
STENOGRAPHERS

	Score in Test VI	Ranking	Rating on Value to Firm by Office Manager
A.....	168	20	20
B.....	163	19	19
C.....	152	18	16
D.....	139	17	17
E.....	134	16	15
F.....	130	15	12
G.....	125	14	14
H.....	118	13	9
I.....	114	12	4
J.....	113	10.5	6
K.....	113	10.5	13
L.....	110	9	18
M.....	105	8	8
N.....	86	7	2
O.....	82	6	7
P.....	78	5	11
Q.....	76	4	10
R.....	72	3	3
S.....	61	2	5
T.....	52	1	1

TABLE II  
RELATION OF SCORE IN TEST VI TO ABILITY IN STENOGRAPHY

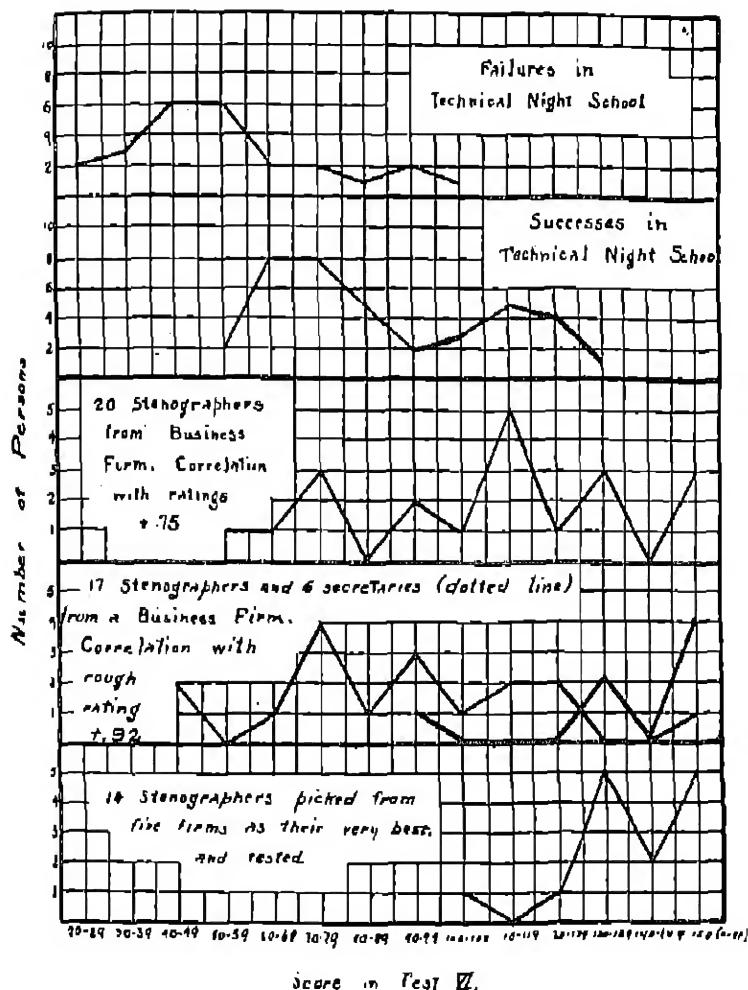
Rating	Score In Bureau Test VI	Position
Good.....	179	Secretary
".....	155	"
".....	152	"
".....	150	"
".....	150	"
".....	139	"
".....	137	"
".....	93	"
".....	155	Stenographer
".....	129	"
".....	125	"
".....	117	"
".....	94	"
".....	91	"
".....	91	"
".....	78	"
"Getting By".....	80	"
".....	72	"
".....	62	"
".....	46	"
Failing.....	74	"
".....	73	"
".....	42	"



# TEST FOR USE IN THE SELECTION OF STENOGRAPHERS 377

CHART II.

Relation of Test VI to Ability in Stenography



## PAPER AND PENCIL RESEARCH IN THE COLLEGE CLASSROOM

By DONALD A. LAIRD, University of Iowa

It was about two decades ago that psychology extended its methods to include more than the prosaic brass instrument laboratory and philosophic convictions. It was about this long ago that the paper and pencil psychologist ushered in the era of tests. Paper and pencil methods were employed in psychological work early in its development. Questionnaires were used to a large extent by the early student of the psychology of religion and during the epoch of child study. The test method, properly termed, however, as it has been developed and brought nearer to perfection by the modern paper and pencil psychologist is of much more recent origin and is rapidly becoming a science unto itself.

More recently the staid study of education has turned to the test methods for a solution of many practical problems in educational procedure in an experimental manner. Still more recently sociology has adopted this method to its subject matter and much is to be expected from its application in this field.

The Great War precipitated the paper and pencil psychologist into a popularity and an esteem scarcely anticipated by the most enthusiastic advocate. In the period of reconstruction a deluge of tests has appeared, well nigh swamping the worker and the student in the onslaught of tests of this and that. The progress of educational tests has been less stimulated by the war impetus. Their development has been more gradual and perhaps more normal but still they are receiving increasing and just recognition and are being turned out in increasing numbers. Truly we are well into the test era.

It is not my purpose in this brief paper to cast a critical eye upon paper and pencil psychology, education, and sociology. Measured in terms of practical results much can be said for them. Viewed from other angles their hues are not nearly so roseate. For the present I must confine my remarks to another aspect of the test method which seems to be largely overlooked.

Under the urge to devise new tests and assist in the standardization of recent tests it has become the quite common

practice for the instructor to use his students not alone for contributing data but to lend a hand in the scoring and tabulating of the results. The pedagogical value of such procedure is at least nebulous if not positively doubtful. Certain it is that those who contribute the data receive negative returns. Especially is this true when, through the kindness of fellow teachers in other departments, a test which pertains not at all to the work at hand is given for the avowed purpose of obtaining data sufficient for the establishment of norms. Such instances are not of infrequent occurrence. Due to the large number of datum necessary for adequate statistical treatment this undesirable aspect becomes much magnified.

Those students who assist in the tabulation and the statistical treatment of the data probably receive some experience which is of slight educational value but whether or not this justifies the exploitation of the masses is an open question which it would seem should be decided in favor of the masses. The principal value attaching to the tabulation and treatment of data by the student is in training for research. Even this can be debated for plainly routine testing is not research nor is it especially conducive to a research attitude. As for the underclassmen—they are simply *Versuchtiere*.

Faculty committees are very prone to discuss ways and means for securing a better use of time by the students. To promote strict class attendance it is decreed that more than three absences from any course will cut into the credits received. Midweek parties are prohibited that the time of the students may not be utilized too largely in extra-curricular activities. It might be well for some of these committees to reflect likewise upon the instructors' opportunities for securing a better use of the students' time in the class room.

Now my thesis—one must always have a thesis nowadays—is that in a rather considerable and rapidly increasing number of instances much of the college and normal school students' time in the classroom is exploited by their teachers in paper and pencil research with but meagre, if any, returns to the student himself. As a judgment of value it would seem that if the students' time and efforts are to be utilized in the classroom it should be in a manner that will bring the students at least some momentary satisfaction if not something of permanent educational value.

I do not think I am making a mountain out of a mole hill. Anyone who has had an opportunity to observe the more recent progress and popularization of the test method cannot have failed to note that this is a factor which is assuming

a large magnitude in proportion to the pedagogical value which attaches to it in classroom routine. Almost innumerable examples could be given of this exploitation of classroom time within a single institution but for reasons which are on the surface it is perhaps best not to be too definite in the present communication. There is also probably little need of dragging in a large array of statistics for every college teacher is or should be aware of the situation which I have brought up for consideration. One who is given to playing with figures might treat this topic with results that would make the infant mortality charts furious with envy.

Probably no one will take offense if I use myself as my "horrible example." I have been guilty of using class time—lots of it—for paper and pencil research in gathering material which had not the remotest connection with their class work. I admit it blandly and complacently for I feel it has the stamp of social approval. The material I obtained, however, turned out to be extremely barren so now we refer to it as preliminary material. (Those who have been reading too much psychoanalysis please note.) To obtain this material three hundred students were detained from more worthwhile class work for a period of fifty minutes. The total time which I thus exploited was fifteen thousand minutes or two hundred and fifty hours. Thus through the willingness of two instructors and the dean of a college who aided and abetted me I consumed more than the equivalent of the time one student would have spent in class work during that quarter.<sup>1</sup>

But my experience is by no means typical. If this preliminary material had warranted more thorough work I would probably not have stopped short of a thousand cases. I also would have used other and more advanced classes for the marking, tabulating, and correlating. Which, again translated into terms of the average students' classroom schedule would mean an entire year's work!

A great deal has been accomplished in a practical way with the use of tests and their development should undoubtedly be furthered. Whether or not the college and university is the proper place for such development, however, is another question and one which I feel should be answered in the negative. The functions of a college or a university are instruction, furthering the sum total of human knowledge, and furthering its own interests. Let us take these up in the reverse order.

---

<sup>1</sup> It is just that I state that this is not the school from which I am writing.

One of the means by which an educational institution may further its own interests is to encourage its staff to contribute from time to time to the sum total of human knowledge. Thus in some respects the last two functions can be combined in one work. With the advent of paper and pencil research *mirabile dictu* it seemed as if by a sudden stroke of providence it would be possible to combine all these functions and fuse them together. Superficially this may be true. The first and all important function of an institution of higher learning, to which all other functions should be contributory, is instruction. And teaching is something far more than using classroom time. It is obvious that the development of tests—educational, psychological, sociological, and what not—in the class period is neither instruction nor education for the average run of underclassmen. The development of such methods and materials in the class room should be discouraged except, perhaps, in special classes organized solely for that purpose.

---

## COMMUNICATIONS AND NOTES

### A NEW LABORATORY IN CHILD PSYCHOLOGY

The Iowa Child Welfare Research Station at the State University has organized a Laboratory in Child Psychology for experimental work with children from 2 to 4 years of age. A new four-roomed building has been constructed and 24 children are now in attendance daily, in two sections from 9 to 12 o'clock. The laboratory is under the immediate direction of Dr. Bird T. Baldwin, Research Professor in Psychology, and Dr. Lorle I. Stecher, Research Assistant Professor, with graduate student assistants.

Oct. 12, 1921

E. T. BALDWIN

## BOOK REVIEWS

*Life—A Study of the Means of Restoring Vital Energy and Prolonging Life.* By Dr. SERGE VORONOFF, Director of Experimental Surgery at the Laboratory of Physiology of the College de France. Translated by Evelyn Roswick Voronoff. Pub. by E. P. Dutton, 1920. 160 p.

The cynic who casts all theories about the prolongation of life into the scrap heap made up of the use of sour milk, tobacco, malt liquors or the abstention from them will not have such ease in discerning flaws in the logic of Dr. Serge Voronoff.

His photographic evidence is hardly debatable, and his vital statistics hardly of the sort which we can suspect of having been juggled to fit his thesis. The book is written in somewhat popular style and the serious student wishes for more detail as to the surgeon's operative method and more exhaustive historical survey of glandular transplantation.

The chief defect of the volume is its omission of references to works of well-known American surgeons who have for at least six years actually performed these operations successfully on human patients. The author's work, however, was carried on chiefly with lower animals such as rams and goats. Nevertheless the book is broad in conception if not wholly so in execution.

In his first chapter he reviews the inverse relation of longevity to organic perfection in the animal series.

The second deals with the true cause of senility, viz. the persistence in our organism of the primitive non-specialized cells incapable of performing the functions of specialized cells, disorganizing them and arresting life. These primitive or conjunctive cells are originally indispensable to the body and we cannot destroy them. Rather must we try to moderate their tendency to rapid multiplication by increasing the vital energy of specialized cells, the sex glands offering the greatest source of such energy. By way of illustration the author cites the mental and physical effects of the removal of these glands in the case of eunuchs, and in contrast the examples of men of genius whose sex functions and artistic productivity were still vigorous very late in life.

Chapter V explains the failure of the Brown-Sequard method of injection in the defective means of preserving the extract, and presents the consequent arguments for the superiority of the graft method. Then follows Voronoff's report to the French Surgical Congress in 1912 of a lamb born of a ewe whose ovaries had been removed and replaced with others. In 1919 he reported grafting testicles upon females giving male characteristics and vice versa, grafting testicles on castrated males and aged males followed by disappearance of their infirmities, restoration of reproductive power and prolongation of life many years beyond that ever hitherto recorded for such animals.

In conclusion he calls attention to the scientific possibility as yet not legally sanctioned, of securing human glands from men killed by accident or from executed criminals, because glands may be kept alive for weeks in refrigerators of zero temperature. But failing this a substitute for the human thyroid is found in that of the higher simian. The identity of secretions and effect of their action on other than the

same species is described. To be specific, a patient whose thyroid mechanism at birth was devoid of blemish had measles at eight years resulting in myxoedema and imbecility. After much experimentation with injection and thyroid tablets, the continuance, suspension and resumption of such medication, a graft from an ape brought a complete recovery and mental awakening. Other interesting examples are reported and finally emphasis laid on the service of augmenting internal secretions of persons the usefulness of whose acquired experience and accumulated knowledge is enhanced by prolonged life and stimulation to intellectual work.

The appendix contains the findings reported to the Biological Society of detailed microscopic examination made by M. Ed Retterer upon Dr. Voronoff's grafts.

*Vassar College.*

MIRIAM C. GOULD.

*A Study of the Mental Life of the Child.* By VON HUG-HELLMUTH, H. Translated by James J. Putnam and Mabel Stevens. Nervous and Mental Disease Pub. Co., 1919. 154 p.

This monograph does not claim to represent the outcome of a judicial weighing of all the factors that go to make the infant and child what they are, says Dr. Putnam in his preface, but it describes a certain very important group of factors that make up the sexuality of childhood.

The author in a way disarms the critic by acknowledging the usefulness of other's modes of classification and choice of emphasis, by recognizing objections to her own plan of treatment and warning that the very special mental tendencies she describes as if existing independently, in sharply defined epochs of childhood in point of fact however, run on through the whole duration of childhood. In no sense is the supposition justified that she conceives of the child's life as a thing marked off into periods on the very arbitrary basis she has adopted for convenience, namely, the nursing period, the play period, and the period of serious study.

Dr. Hellmuth writes in a highly popular style, using for illustrative material her own nephew and many observations of Preyer, Shinn, Scupin and Sully, who have studied the child by the biographical method but not the Freudian interpretation. Stern and G. Stanley Hall are additional sources, but there is little or no reference to further contemporary investigation.

Whatever our views concerning the acceptability of the Freudian accent on the entire gamut of situations and response in childhood, no reader can fail to grant the universality of most of the observed behaviour on which she comments. We are not so ready to admit all of the author's generalizations whose logic is glaringly open to attack by experimentalists who have contradictory data.

*Child behaviour. A critical and experimental study of young children by the method of conditioned reflexes.* By FLORENCE MATEER. Badger, 1918. 239 p.

This is a technical volume dealing with the Pavlov Method which was first applied to normal children by Krasnogorski in studying the development and decadence of conditioned reflexes and inhibitions in sucklings and young children, later to cases of neuropathic and feeble-minded children.

The author criticizes Krasnorgorski's account of his experiments for its serious omissions and says "the claims made by Krasnorgorski for his method and findings are so great that a verification or disproof of them seems to be demanded from the standpoint of both experimental and clinical psychology. If the method is so easily and accurately applicable and allows so many variations, it opens up to the experimentalist a most promising avenue of attack upon the infant mind. If it differentiates with such accuracy the defective and pathogenic child while he is still young enough for preventive education and corrective therapy to have a chance it will revolutionize our present clinical methods." It is this that Dr. Mateer seeks to ascertain by a modified methodology and technique. She describes in detail her procedure of procuring the 50 unselected normal children under 7 years of age in a village, the establishment of rapport, the activities preliminary to the experimentation, the social, hereditary physiological study of the subjects, the other groups of defective children in the Mass. State School for Feeble-minded and of suspected or observation cases similarly tested and the additional extensive studies she made of individual children.

Preliminary experiments upon a few subjects were made with the aim of studying and understanding the method itself. A series of standard clinic tests for determination of mental levels were used such as the Goddard Revision of the Binet Scale, the Yerkes Point Scale, Healy Boards, the Goddard adaptation and form boards. Kymograph records were then made of movements in establishing conditioned opening of the mouth through a cutaneous stimulation of the inner side of the right arm above the elbow, originally associated with feeding sweetened water or candy. Specificity and discrimination of cutaneous stimuli were studied in the same manner.

The main experimental results are as follows: (1) 3-9 trials necessary in normal children to form a sensory-motor association, 3-18 in defectives. The necessary number decreases as age increases up to 60 months. Under two years boys are more rapid than girls, above two years the order is reversed. The number of trials varies more among girls than boys. The rapidity of learning or acquiring a conditioned reflex correlates with intelligence.

(2) The number of trials necessary to obtain memorial refunctioning after an interval of 24 hours is from 2-7. 70% of the normal subjects reacted at once; 57% of defectives with 2 trials, the rest 3-7.

(3) The most rapid learners retain best. Girls are the most rapid.

(4) The number of trials necessary to effect unlearning of an associative act when conditioning stimulus is absent is 3-12 for normals, 8-21 for defectives. The number decreases as age increases, and number of trials for learning increases. Boys are most rapid.

(5) No perceptible difference in children of all ages or either sex in number of trials necessary to correct the unlearning and restore refunctioning of first associations.

The author makes no claim of finality in her finding, but points out their significance in our progress towards the use of objective methods of child study.

She does not advocate their substitution for other clinical methods, but simply their addition because our information concerning the organism becomes more certain by studying the variations of behaviour under definitely ascertained conditions of stimulation.

To quote Miss Mateer "Psychogenetic or biographical studies were and are valuable. They have been the pioneer path-breakers that



indicate the extent of the problem confronting us, but now exact definitization of facts demands quantitative studies under controlled experimental conditions and an unbiased presentation of the results obtained." However she warns against inferring too much from arbitrary situations, cold manipulation, and forgetfulness of the child as an individual.

Her survey of the history and results of child study up to the present is thorough, yet admirably concise.

*Vassar College*

MIRIAM C. GOULD.

*A Study of Women Delinquents in New York State.* By MABEL R. FERNALD, MARY H. S. HAYES and ALMENA DOWLEY, with a statistical chapter by Beardsley Ruml and a preface by Katherine Bement Davis. Publication of the Bureau of Social Hygiene, N. Y. The Century Co., 1920. 542 p.

This is the first extensive study of women delinquents given expert statistical treatment with modern methods. Its 225 tables are critically analyzed and conclusions summarized at the end of each chapter, making the work satisfying alike to readers with or without a thorough comprehension of the technique used.

Unfortunately the data are confined to mental capacity and the main facts, of personal and environmental history, because adequate data for physical and medical facts could not be obtained.

The groups studied were (1) N. Y. State Reformatory women between the ages of 16 and 30 convicted of felonies and misdemeanors; (2) Auburn State Prison felons; (3) Magdalen Home Sex offenders; (4) N. Y. County Penitentiary; (5) Workhouse, and (6) Women's Night Court cases. Inebriates were omitted because of inability to obtain accurate information from them, but this eliminates many mental defectives and therefore introduces a serious source of error. A further defect is the variation in time period over which cases in the different institutions were examined. No control series for comparison with the normal population was used and comparison with men criminals was futile because the offenses of male convicts are not comparable.

Mental condition was ascertained by the Binet Simon 1911, Yerkes, Bridges and Stanford Revision Scales, the Wooley series, some educational and individual standard tests of ability. Social investigations were made through visits to the home, relatives, employees and other agencies having knowledge of the offender. A correlation of .38 was found between delinquency and intelligence, also significant correlations between mental capacity and type of occupation, mental capacity and earning capacity, age and number of previous convictions, age at first conviction and condition of the home.

No correlation was found between age of leaving school and age of first conviction, nor between grade reached in school and number of convictions. Lack of education, therefore, is not responsible for delinquency.

Furthermore, the popular conception that low and irregular wages are responsible finds no confirmation here, for a positive correlation could not be found between earning capacity and number of convictions.

Foreign white women made up a smaller percentage in each institutional group than is present in the population as a whole, yet the offenses of the foreign delinquents were more serious on the whole than those of the natives.

There were more domestic workers in the group than in the general population and a lower average standard of education among them than in the other occupations represented. The older women had a lower grade of education than that reached by the younger cases.

No relation appeared between intelligence and habitual use of alcohol, drugs or cigarettes, but the prostitutes were distinctly mentally inferior to those whose crimes were against property.

In 15.9% of the delinquent's families some other member had been convicted of crime, but the data was too inadequate to enter into a study of hereditary aspects. Most of them were from self-supporting families and born of parents not at age extremes.

In several particulars the findings are in substantial agreement with those of Goring's figures for the English convict, but the authors are cautious in their deductions and fully aware of the inconclusiveness of their evidence.

MIRIAM C. GOULD.

*Vassar College.*

# INDEX OF SUBJECTS

(The subject headings of all contributions except Book Reviews and Notes are printed in SMALL CAPITALS.)

ADVERTISING, MINOR STUDIES IN THE PSYCHOLOGY OF	5	GENERAL INTELLIGENCE	72
AGGRESSIVENESS, THE MEAS- UREMENT OF	97	GIFTED JUVENILE AUTHOR, PRELIMINARY REPORT ON	163
APPLIED PSYCHOLOGY IN THE UNITED STATES, STATUS OF	1	I. Q., TWO CASES SHOWING MARKED CHANGE IN,	156
BEDAUX UNIT PRINCIPLE OF INDUSTRIAL MEASUREMENT	119	INDIANS, TESTS ON FULL AND MIXED BLOOD	359
CHARACTER TRAITS, EXPERI- MENTAL STUDY OF	297	INCOMPLETE ORDER OF MERIT RATINGS, A STATISTICAL METHOD FOR	261
Child Behavior	383	INDUSTRIAL MEASUREMENT, THE BEDAUX UNIT PRIN- CIPLE OF,	119
CHILDREN COMPARED BY GOD- DARD FORM BOARD	253	INDUSTRIAL PSYCHOLOGY, 127; IN EUROPE,	287
CHINESE CHILDREN IN SAN FRANCISCO, INTELLIGENCE OF	267	INTELLIGENCE, ESTIMATED BY MEANS OF PRINTED PHOTO- GRAPHS, 152; OF CHINESE CHILDREN IN SAN FRAN- CISCO, 267; TESTS,	184
COMPTOMETER OPERATORS, SE- LECTION OF	275	JOURNAL OF PRACTICAL PSY- CHOLOGY. TRANSLATION OF INTRODUCTORY ARTICLE	85
CONTROL OF CONDUCT BY SUG- GESTION	26	JOURNALISTIC APTITUDE, TEST SERIES FOR	46
CORRELATIONS BETWEEN ESTI- MATES OF VOLITIONAL TRAITS AND THE RESULTS FROM THE DOWNEY "WILL-PROFILE"	159	Lehrbuch der experimentellen Psychologie	187
Courtis Standard Practice Tests in Handwriting	92	Life—A Study of the Means of Restoring Vital Energy and Prolonging Life	382
DEARNORN TESTS IN RELATION TO STANFORD	284	Manhood of Humanity	189
DRAFTSMEN, EFFECT OF MUSIC UPON	350	Mental Life of a Child	383
EDUCATIONAL GUIDANCE AND TESTS IN COLLEGE	32	Methods and Results of Test- ing School Children	190
ESTIMATING INTELLIGENCE BY MEANS OF PRINTED PHOTO- GRAPHS	152	MOVING PICTURES, SCHOOL RATINGS AND	284
FEELLEMINDED, THE RESULTS OF REPEATED MENTAL RE-EX- AMINATIONS OF	195	MUSIC, EFFECT ON DRAFTS- MEN	350
Foundation of Spiritualism	91	Mysticism, Freudianism and Scientific Psychology	90
FRESHMEN, NORMS AMONG UNIVERSITY	318	NEGRO, WHITE AND INDIAN WORK CURVES	14
Fundamentals of High School Mathematics	188		

- NOTES ON FUNDS FOR SCIENTIFIC RESEARCH, 95; INDUSTRIAL RESEARCH LABORATORIES, 192; INDUSTRIAL PSYCHOLOGY IN EUROPE, 287; LABORATORY IN CHILD PSYCHOLOGY, IOWA STATE UNIVERSITY, 381; NATIONAL INSTITUTE OF INDUSTRIAL PSYCHOLOGY, 290; PSYCHOLOGICAL WORK IN THE UNITED STATES ARMY 192
- PAPER AND PENCIL RESEARCH IN THE COLLEGE CLASSROOM 378
- Personality Culture by College Faculties 94
- Personnel Administration 92
- PHOTOGRAPHS, ESTIMATING INTELLIGENCE BY MEANS OF 152
- PICTORIAL COMPLETION TESTS II 225
- PRACTICAL PSYCHOLOGY, A NEW JOURNAL OF 85
- PREDICTIVE VALUE OF SHORT INTELLIGENCE TESTS 184
- Psychopathology 292
- RACE NORMS, NEED FOR CAUTION IN ESTABLISHING 179
- REPEATED MENTAL RE-EXAMINATIONS OF FEEDLEMINDED 195
- SCHOOL RATINGS AND MOVING PICTURES 284
- SEX DIFFERENCES, FURTHER DATA 78
- SPECIAL ABILITY TESTS 127
- STANFORD TESTS IN RELATION TO DEARBORN 284
- STATISTICAL METHOD FOR INCOMPLETE ORDER OF MERIT RATINGS 261
- STENOGRAPHERS, METHODS FOR SELECTION OF, 275; TESTS FOR THE SELECTION OF 373
- STUTTERING, ANALYSIS OF THE CAUSES OF 340
- SUGGESTION, CONTROL OF CONDUCT BY 26
- TEACHING PERSONALITY QUOTIENT 39
- TESTS FOR THE DISCOVERY OF VOCATIONAL FITNESS, 240; IN INDUSTRY, 57; FOR JOURNALISTIC APTITUDE, 46; FOR SPECIAL ABILITIES, 127; PICTORIAL COMPLETION, 225; PREDICTIVE VALUE OF, 184; FOR USE IN THE SELECTION OF STENOGRAPHERS 373
- Trade Tests 89
- UNSELECTED GROUPS, PROBLEM OF 64
- WHITE, INDIAN AND NEGRO WORK CURVES 14
- Women Delinquents in New York State 385

# INDEX OF AUTHORS

(The names of authors of original contributions are printed in SMALL CAPITALS.)

ANDERSON, L. DEWEY	152	KITSON, HARRY D.	5, 287
ANDERSON, LEWIS O.	340	Korzybski, Alfred	189
ARLITT, ADA HART	179	KUHLMANN, F.	195
BÉDAUX, CHARLES E.	119	LAIRD, DONALD A.	378
BERG, DAVID E.	94	LEONARD, CLINTON S.	91
BILLS, M. A.	275, 373	LINK, HENRY C.	85
BREGMAN, ELSIE OSCHRIN	127	Mateer, Florence	383
BRUDACHIER, A. R.	39	MEENES, M.	92, 190, 191
Chapman, J. Crosby	89	Metcalf, Henry C.	92
Child, Emily	190	MYERS, GARRY C.	26
Clark John Roscoe	188	MOORE, H. T.	97
COLVIN, STEPHEN S.	32	PRATT, C. C.	188
Courtis, Stuart A.	92	PRESSEY, L. W.	78
Dewey, Evelyn	190	PRESSEY, S. L.	64, 72, 78
Dowley, Almena	385	REAM, M. J.	184, 261
Dunlap, Knight	90	Root, W. T.	156
FENTON, JESSIE C.	163	RUCH, G. M.	159
Fernald, Mabel R.	385	Rugg, Harold Ordway	188
FILTER, RAYMOND O.	297	Ruml, Beardsley	190
FREYD, MAX	46	SCHWENNING, G. T.	90, 94
Fröbes, J.	187	Shaw, Lena A.	92
GARTH, THOMAS R.	14, 359	Smith, W. Whately	91
GATEWOOD, ESTHER L.	350	Tead, Ordway	92
GAW, FRANCES I.	384	TERMAN, LEWIS M.	1, 163
GILLILAND, A. R.	97	VITELES, MORRIS S.	57
GOULD, MIRIAM C.	383, 385, 386	Von Hug-Hellmuth, H.	383
HANSEN, C. F.	184	Voronoff, Serge	382
Hayes, Mary H. S.	385	WATTS, FRANK	240
HEALY, WM.	225	WHITCHURCH, ANNA KELL-	
HERRICK, D. S.	253	MAN	318
Kempf, Edward J.	292		